

UNIVERSITY OF COLOGNE

DOCTORAL THESIS

Human Perspectives on the Use of Color in Data Visualization

Author:
Laura Pelchmann

Dissertation for the degree of
Doctor of Natural Sciences (Dr. rer. nat.)
in the
Department of Mathematics and Computer Science



2026

"Girls are capable of doing everything men are capable of doing. Sometimes they have more imagination than men."

Katherine Johnson

Abstract

The discriminability of color differences is a critical yet under investigated factor in the effectiveness of node-link diagrams. This work consolidates five versatile investigations that examine how various visual encodings affect viewers' ability to detect color differences in node-link diagrams and scatterplots. These visual encodings include link hues, node outlines, hue-luminance color schemes for large value ranges, semantically resonant multi-label color schemes, and factors affecting to detect qualitative differences and structural changes across graphs.

Across several online experiments and controlled laboratory studies, we varied link coloration (complementary versus hue-matched), node-outline direction, outline color, and outline widths, and measured participants' accuracy in identifying color differences between nodes. Consistently, complementary-colored links amplified node-color discriminability irrespective of underlying topology, whereas hue-similar links degraded it, while outward directed, thin black or yellow/orange outlines further enhanced discriminability for blue nodes.

In a separate optimization study, we extended the Order of Magnitude Colors (OMC) scheme by systematically selecting hues and adapting luminance. Thereby we ensured color deficiency robustness, and scaling to arbitrary exponent counts. A quantitative color-metric analysis and a human-subject evaluation demonstrate that two optimized versions of the well known Glasgow scheme significantly outperform the original OMC for visualizing large value ranges (LVRs).

A further investigation addresses the generation of semantically resonant color schemes for multi-label classes. Through two user studies and a domain validation with meteorologists, we derive three effective design strategies and implement an algorithm that produces multi-label color schemes, confirming its utility across diverse real-world datasets.

Finally, a comparative study of directed acyclic graphs (DAGs) reveals that node and edge coloring, together with strategically placed uniform color regions, markedly improves the detection of structural differences across graphs under time pressure.

Collectively, these findings advance the understanding of perceptual factors on color discriminability in node-link diagrams and scatterplots and inform the design of more accurate, accessible, and task-effective visualizations.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Use of Color in Data Visualization for Quantitative Data	7
1.2.1	Background and Related Work	7
1.2.2	Research Questions	7
1.2.3	Contributions	8
1.3	Use of Color in Data Visualization for Qualitative Data	9
1.3.1	Background and Related Work	9
1.3.2	Research Questions	10
1.3.3	Contributions	10
1.4	Outline of the Thesis Structure	11
2	Human Perspectives on the Use of Color in Data Visualization for Quantitative Data	13
2.1	Discriminability of Node Colors in Node-Link Diagrams: Link Hues	13
2.1.1	Introduction	13
2.1.2	Just Noticeable Differences	13
2.1.3	Discriminability of Node Colors in Node-Link Diagrams and the Influence of Link Colors	14
2.2	Discriminability of Node Colors in Node-Link Diagrams: Node Outlines	35
2.2.1	Introduction	35
2.2.2	Visual Highlighting	35
2.2.3	Discriminability of Node Colors in Node-Link Diagrams and the Influence of Node Outlines	36
2.3	Discriminability of Node Colors: Large Value Ranges	53
2.3.1	Introduction	53
2.3.2	Large Value Ranges	53
2.3.3	Color Vision Deficiencies	53
2.3.4	Effective Color Schemes for Large Value Ranges Optimizing Color-Coding Criteria	54
3	Human Perspectives on the Use of Color in Data Visualization for Qualitative Data	77
3.1	Discriminability of Node Colors: Semantic Color Associations	77
3.1.1	Introduction	77
3.1.2	Human Strategies for Semantic-Based Colors Representing Multi-Label Classes	77

3.2	Discriminability of Node Colors in Node-Link Diagrams:	
	Graph Comparison	95
3.2.1	Introduction	95
3.2.2	Factors Influencing Visual Comparison of Colored Directed	
	Acyclic Graphs	95
4	Conclusion and Future Directions	109
4.1	Conclusion on Human Perspectives on the Use of Color in Data Visualiza-	
	tion for Quantitative Data	109
4.2	Conclusion on Human Perspectives on the Use of Color in Data Visualiza-	
	tion for Qualitative Data	112

Chapter 1

Introduction

1.1 Motivation

Color serves as an important visual variable for information visualization, allowing for a fast and easy differentiation of the data attribute values they represent. The human brain can preattentively process color, which helps us to quickly identify data structures and anomalies [8, 168]. This is advantageous in various applications, especially when data e.g., medical data (e.g., infection risks) or meteorological data (e.g., ice water content) needs to be communicated to non-expert viewers.

Color as visual variable is particularly important in graph visualization, where positional encoding of both nodes and edges, is constrained by the data as well as the chosen layout algorithm. Similarly, in scatterplots, the positions of data points are determined by the data itself. Consequently, color plays a central role for the visual encoding of further data attribute values in both types of visualizations [26, 55, 95, 127, 147, 167].

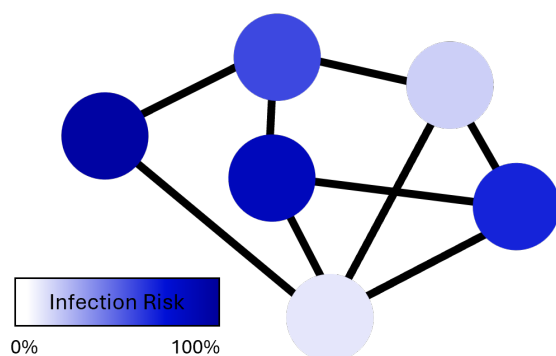


FIGURE 1.1: Example visualization of a contact network visualizing the “Infection Risk” using node colors.

Colors are often used to specifically represent attribute values within nodes, illustrating both quantitative and qualitative values. To ensure that viewers can easily and accurately interpret these visualized values, for example, to correctly identify chains of infection (see Figure 1.1), financial transactions, or cluster of interest in power grids, it is important that the viewers can distinguish between the colors representing different attribute values. Differentiable colors are therefore an essential component of effective data visualization, as they directly impact the correct and accurate interpretation of the visualized data. [86, 96, 152, 174].

Ensuring color discriminability in visualizations is a challenge, as multiple nodes or points, and therefore their colors, must remain discriminable when viewed simultaneously. A starting point is to consider pairwise color discriminability, i.e., every possible pair of nodes and circles in the diagram should be reliably discriminable from one another. If individual pairs of colors cannot be distinguished, users may easily confuse

the corresponding node attribute values when they appear side by side in a visualization. While pairwise discriminability alone does not guarantee discriminability within the context of an entire diagram, because additional factors such as spatial arrangement, visual clutter, and context-dependent interactions can influence perception, it is a necessary prerequisite for creating effective color palettes [43, 168].

In recent years, various studies have been conducted to investigate the discriminability of colored circles in scatterplots for different types of data. For example, these studies have examined both the optimization of certain categorical color schemes for adjacent clusters [167] and the minimum color differences required between adjacent circles intended to represent quantitative data attribute values [152]. Such studies are largely absent for node-link diagrams. The problem, however, is that despite the visual similarity, findings from the scatterplot studies cannot simply be transferred to node-link diagrams. That is because the relational data aspect of the attribute values is conveyed via edges, that can visually interfere with the discriminability of the node colors (see Figure 1.2, the blue edge connects two differently colored nodes, impeding with the discriminability). Further studies that are comparable to the existing studies investigating scatterplots are therefore needed, in order to better assess whether the findings on color discriminability in scatterplots can be applied to the discriminability of node colors in node-link diagrams, or whether different color differences are required to truthfully represent node attribute values.

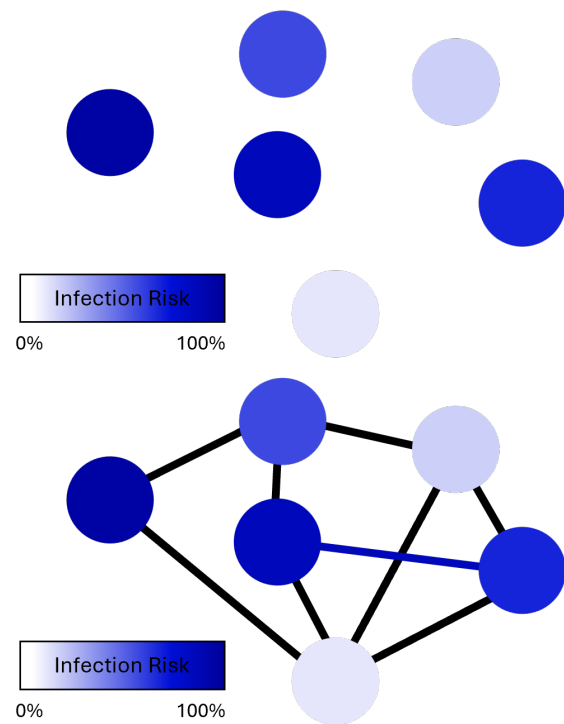


FIGURE 1.2: Example visualization representing the “Infection Risk” using node colors. The node attribute values are the same in both visualizations, no connections between points (top), while using colored edges to represent potentially important interactions between people (bottom).

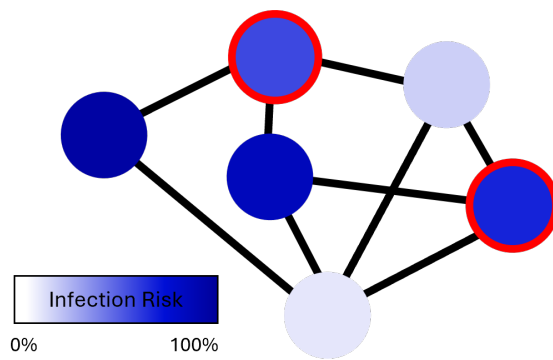


FIGURE 1.3: Example visualization of a contact network that visualizes the “Infection Risk” using node colors highlighting the risk of two individuals.

The influence of edges on the discriminability of node colors shows that additional visual elements can affect the discriminability of node colors in node-link diagrams and therefore need further investigations. For example, other visual elements that can be added to node-link diagrams are node outlines, which can be used to highlight particularly relevant nodes and their underlying attribute values for example infected persons in a medical contact network (see Figure 1.3). These outlines are typically placed around the nodes to draw attention to selected nodes while preserving the nodes’ original color coding and therefore the at-

tribute values being represented. However, by introducing an additional color in the immediate vicinity of nodes, outlines can also alter the discriminability of node colors. Therefore, it is important to understand how different outline designs influence color discriminability to ensure that highlighting techniques improve visual orientation without interfering with the interpretation of color-coded information. While research on this aspect of color discriminability has already been conducted for scatterplots [96], the question remains largely open for node-link diagrams.

While data such as infection risk typically represent a limited range of values, there are quantitative datasets in many fields, particularly in medicine and biology, but also in meteorology and finance, that cover significantly larger ranges of values. When visualizing those so called large value ranges, simple linear color schemes often do not provide enough discriminable colors to cover the entire range. Depending on the virus, for example, infected individuals may have between 20 and 10,000,000 viral copies per mL of blood during the course of an illness. Mapping such a wide range of values to a limited color range can therefore compress the discriminability in certain areas, which negatively affects the readability of the data. The challenge, then, is to display this data in a way that allows for clear differentiation (see Figure 1.4). In this case, we must ensure not only that the colors are discriminable along a single gradient but also that they are discriminable from one another across different gradients, especially when aiming to be inclusive for people with color vision deficits.

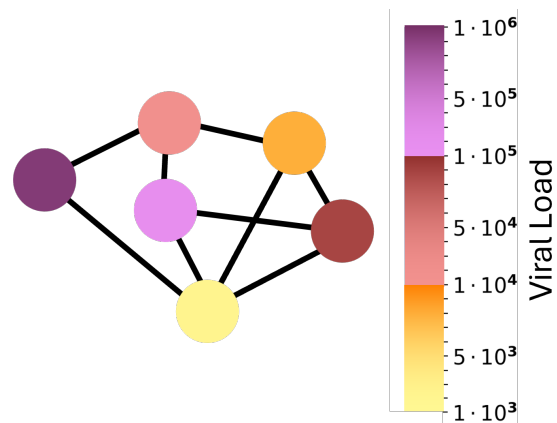


FIGURE 1.4: Example visualization of a contact network that visualizes the “Viral Load” using node colors through a color scheme for large value ranges.

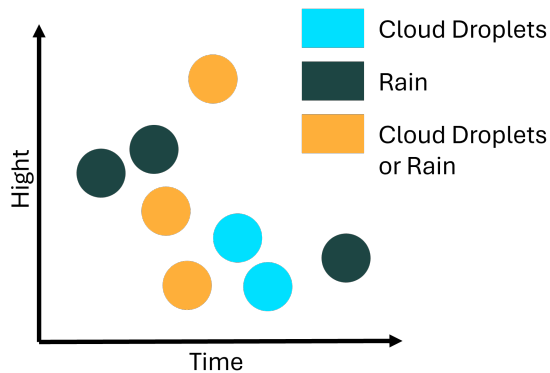


FIGURE 1.5: Example visualization of a time height plot representing exemplar meteorological particle measurements.

shown in shades of blue. However, it is not always possible to follow all semantic color associations from a dataset. Conflicts can occur, for example, when different categories within a dataset have similar color associations (such as in meteorological data, where the color associations for clouds, rain, hail, fog, and drizzle include many shades of blue and gray) or when data categories, due to measurement inaccuracies, cannot be clearly distinguished from one another (see Figure 1.5). In such cases, using colors based on semantic associations may conflict with the choice of easily discriminable colors.

While semantic color schemes facilitate the interpretation of categorical data within a single visualization, other challenges arise when viewers are comparing multiple visualizations. In comparative analyses, users often need to identify changes in data structures through color changes between corresponding nodes or points, rather than interpreting individual colors within a single diagram (see Figure 1.6). Consequently, the effectiveness of the visualization depends not only on the semantic meaning of the chosen colors, but also on how easily differences of colors and their attribute values can be detected across diagrams. Understanding the factors that influence the detection of these cross-diagram color differences is therefore essential for designing visualizations intended to support accurate comparison tasks.

When visualizing qualitative data, perceptual discriminability is not the only aspect to consider. In contrast to the visualization of quantitative data, where colors represent numerical values, qualitative data with real-life context can often be divided into discrete data categories that benefit from color assignments and convey the semantic meaning of these categories. By assigning semantically resonant color associations to these categories, semantic color schemes can enable faster interpretation while ensuring that the categories remain discriminable. When displaying weather data, for example, a sunny day is typically shown in shades of yellow, while a rainy day is

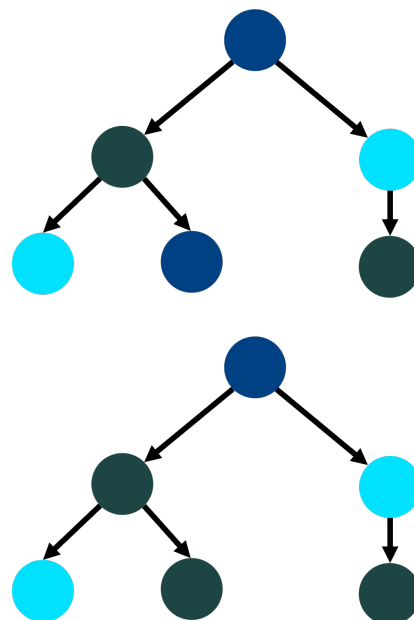


FIGURE 1.6: Example visualization of two node-link diagrams representing the change of data attribute values between them.

The studies presented in this thesis address the previously described challenges that influence color discriminability in information visualizations. They range from perceptual limits of discriminability in quantitative color schemes and large value ranges over visual significance of semantic color encoding, to the comparison of colors across node-link diagrams. Although these challenges might arise in different areas of application, they share the goal of enabling viewers to truthfully distinguish color-coded data attribute values. The overarching goal of this work is therefore to develop design recommendations that improve color discriminability across various visualization tasks. By transforming the findings from these studies into actionable guidelines, this work aims to support both inexperienced users in selecting effective colors for nodes and points as well as their surrounding primitives and experienced visualization designers in making informed design choices when creating or evaluating colored node-link diagrams and scatterplots (see [Figure 1.7](#)).

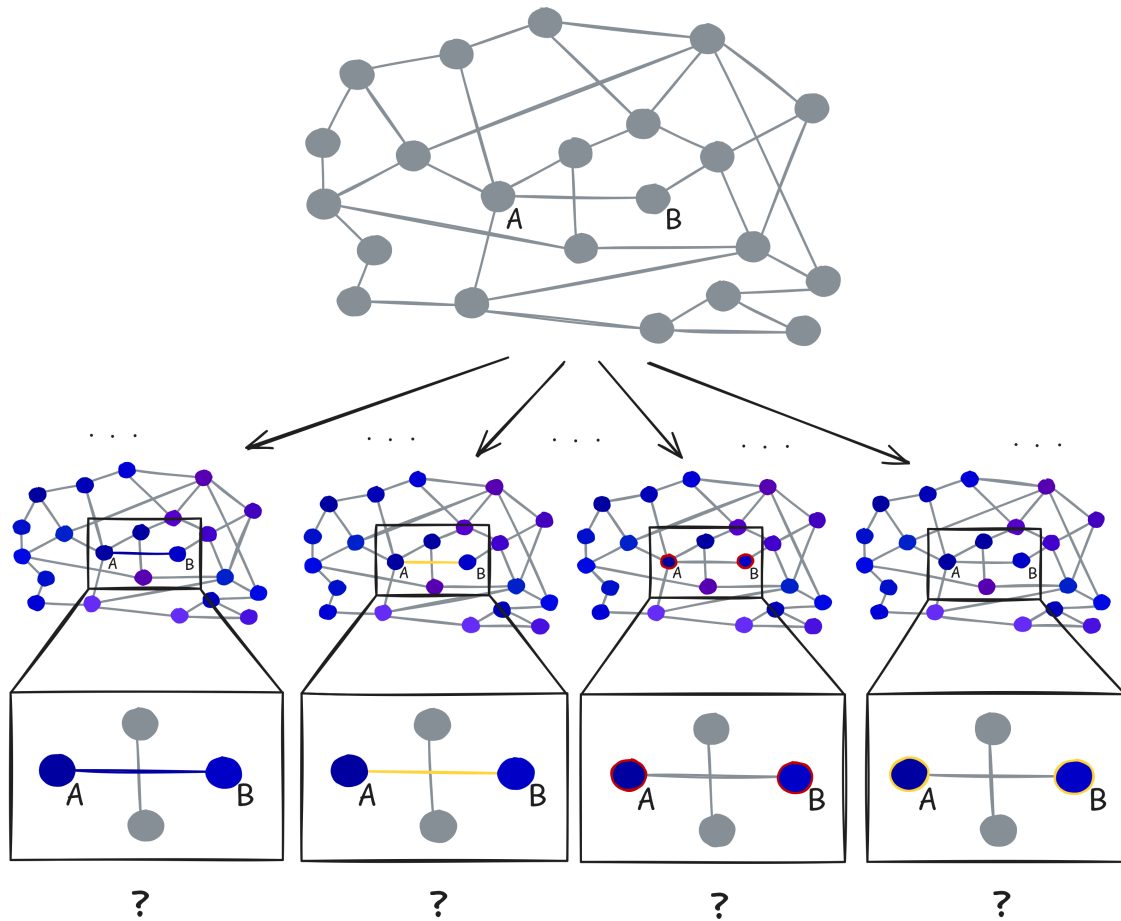


FIGURE 1.7: The goal of this thesis is to support users in choosing suitable colors for creating visualizations to ensure the best possible color discriminability. In this example the color differences between nodes A and B are the same across all four node-link diagrams. With this work we want to investigate which of these solutions should be applied in order to enhance the discriminability between the two nodes.

To provide a better context, we have divided this work into two parts based on the two different types of data input: quantitative data and qualitative data.

The first topic – **Human Perspectives on the Use of Color in Data Visualization for Quantitative Data** – focuses on the representation of quantitative data attribute values as node and point color. The challenge lies in ensuring that different value ranges are represented by the gradient of a single hue, while maintaining the recognizability of value differences within the diagram, even when those differences are only incremental.

The second topic – **Human Perspectives on the Use of Color in Data Visualization for Qualitative Data** – focuses on the representation of qualitative data attribute values as node and point colors, with the challenge that human perception can reliably distinguish only a very small number of colors. Datasets with larger quantities of categories to be displayed or categories to be displayed with strong semantic color associations can therefore present problems for a clear color distinctions for both within and between diagrams.

Together, the two sections of this dissertation contribute to a deeper understanding of the factors that influence the discriminability of color differences in node-link diagrams and scatterplots. Furthermore, we derive guidelines for which color combinations should be used when creating node-link diagrams and scatterplots to best support discriminability and, thereby, of the underlying data attribute values.

1.2 Use of Color in Data Visualization for Quantitative Data

1.2.1 Background and Related Work

The field of color perception identifies three distinct aspects of color: hue, lightness, and saturation [23]. These three dimensions of color impact the discriminability of colors differently. Research shows that people interpret quantitative differences more accurately through changes in lightness than through changes in hue [23, 168]. Hence, lightness is deemed the most critical channel for encoding quantitative data attribute values, usually representing higher values through darker colors. In doing so, equal quantitative differences between data attribute values should appear as equal perceptual differences in the corresponding colors, ensuring that viewers do not overemphasize some value differences while overlooking others. Color schemes that do not follow this perceptual uniformity can distort the understanding of data because certain transitions become visually exaggerated [26].

The challenge for color schemes defined solely by lightness gradients lies with the limited range of color changes. While the whole data range needs to be represented, even small differences in the data must be visible through correspondingly different colors. Consequently, these color differences must be as small as possible to cover a potentially large data range with the color scheme. At the same time, they must be large enough to be perceived by human viewers [152].

The aim of the following works is to identify factors that influence the perception of these required minimal color differences and, in doing so, to formulate guidelines that assist with the color selection when creating node-link diagrams. Specifically, the included papers investigate the impact of different node hues, edge topologies, edge hues, and node outline colors, width and directions. Furthermore, we modify the Orders of Magnitude (OMC) color scheme for the visual representation of large data ranges to ensure accurate color discriminability, even for individuals with potential color vision impairments, following up on the work of Braun et al. [21].

1.2.2 Research Questions

To pursue these goals, this work answers the following research questions:

- **RQ1.1:** How should edge and node colors be chosen to ensure the discriminability between node colors that represent quantitative graph data in contrast to scatterplots?
- **RQ1.2:** How should node outlines be designed to ensure the discriminability between node colors that represent quantitative graph data in contrast to scatterplots?
- **RQ1.3:** Which color scheme displays color differences for quantitative data with large value ranges in data visualization accurately for people with color vision deficiencies?

Understanding the impact of surrounding elements in node-link diagrams on the discriminability of color differences between nodes can enhance the interpretability of the data presented. Therefore, the answers to these research questions represent important advances toward the effective color representation of data attribute values in nodes and, consequently, toward effective color designs for node-link diagrams.

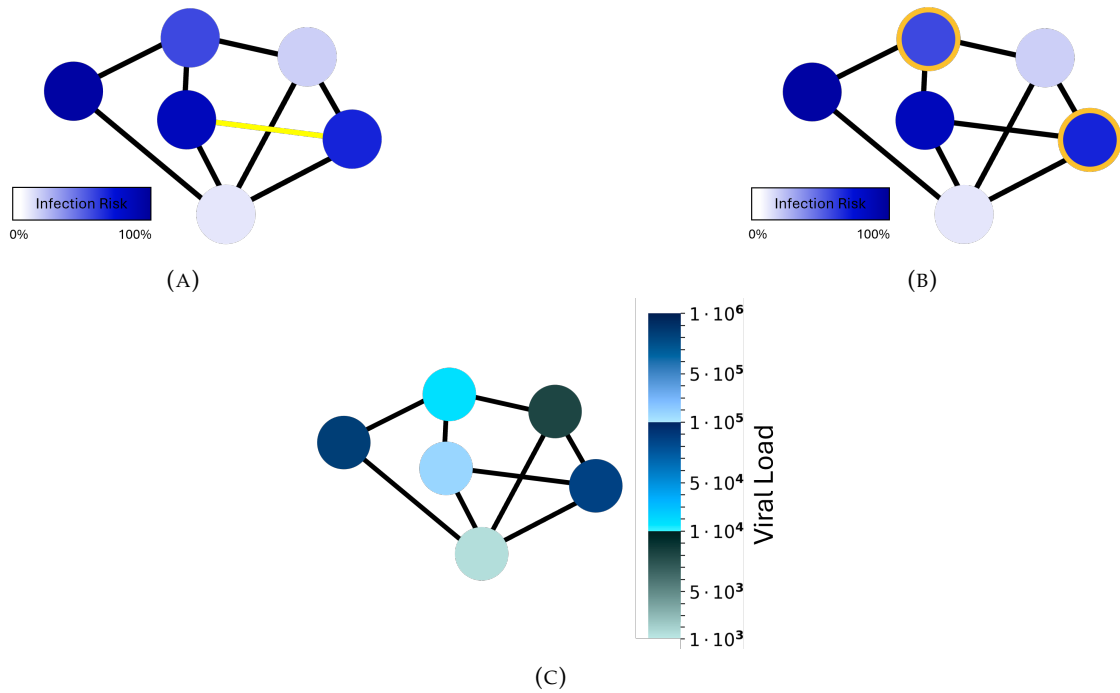


FIGURE 1.8: Overview of the research contributions on human perspectives on the use of color in data visualization for quantitative data.

1.2.3 Contributions

More specifically, these papers make the following contributions:

- Guidelines for the use of color in node-link diagrams to enhance the discriminability of node colors in small graphs with quantitative node encoding. These guidelines apply to color selections for both nodes and their associated edges (see [Figure 1.8a](#)).
- Guidelines for node outlines in node-link diagrams to highlight nodes without distorting the underlying quantitative color encoding. These guidelines apply to color, width and direction of the node outlines (see [Figure 1.8b](#)).
- Two improved color schemes for visualizing large value ranges in various application domains. We address the limitations of the previously developed Order of Magnitude Color scheme in terms of hue selection, luminance adaptation, color-deficiency robustness, and adaptability to various number of exponents (see [Figure 1.8c](#)).

Importantly, all findings and developed color schemes are domain-agnostic and are applicable to any context. They are intended to provide a basis for future research that is urgently required in the domain of color perception in node-link diagrams and scatter-plots.

1.3 Use of Color in Data Visualization for Qualitative Data

1.3.1 Background and Related Work

Categorical color schemes are used to represent qualitative data attribute values, that do not contain any kind of ordering [24, 57]. Through such schemes, the data categories are primarily distinguished through variations in hue while maintaining similar levels of lightness and saturation to avoid implying visual hierarchies [24, 168].

However, a high number of categories can lead to difficulties in color discrimination, as the perceptual distance between hues decreases and users may thus struggle to discriminate colors that are too similar [59, 168].

This problem is further complicated by color vision deficiencies, which can make certain hue combinations difficult or impossible to discriminate for some viewers [57]. In addition, contextual effects and interactions between neighboring colors can influence color perception and reduce discriminability [168]. Consequently, categorical color palettes should be designed to maximize perceptual differences between hues and to avoid color combinations that are frequently confused [57, 59].

Research has shown that the number of well discriminable colors representing different categories is limited, therefore making hue selections an essential aspect of designing qualitative data visualizations [24, 59]. While the ColorBrewer Tool [57] offers categorical color schemes with up to 12 different hues, Healy et al. [59] conducted a series of visual search studies and found that only up to seven different colors could still be identified rapidly and accurately, whereas larger color sets resulted in substantially reduced performance by their participants.

In the case of qualitative datasets, there are often semantic color associations for real-life data categories. Semantic color schemes aim to assign colors based on these associations, rather than using randomly selected color mappings. By following familiar word-color combinations, such as blue for water or red for danger, these schemes can reduce the cognitive effort needed to understand a visualization [93]. Research on word-color associations has shown that many associations are consistent across viewers, providing a basis for designing semantically meaningful color schemes [81] and that effective color mappings depend not only on perceptual discriminability but also on semantic discriminability. This refers to the extent to which users can clearly recognize the intended connection between colors and categories [142]. Therefore, semantic color schemes need to align with viewers' expectations while considering possible cultural differences that may influence the color's meaning [2].

The aim of this thesis is to identify strategies for designing color schemes for data sets whose classes do not consist of distinct categories, but to which multiple labels may be assigned under certain circumstances, thereby complicating the assignment of semantic colors. In addition, this thesis seeks to identify surrounding factors in graphs that influence the recognition of color differences representing quantitative data attribute values across those graphs.

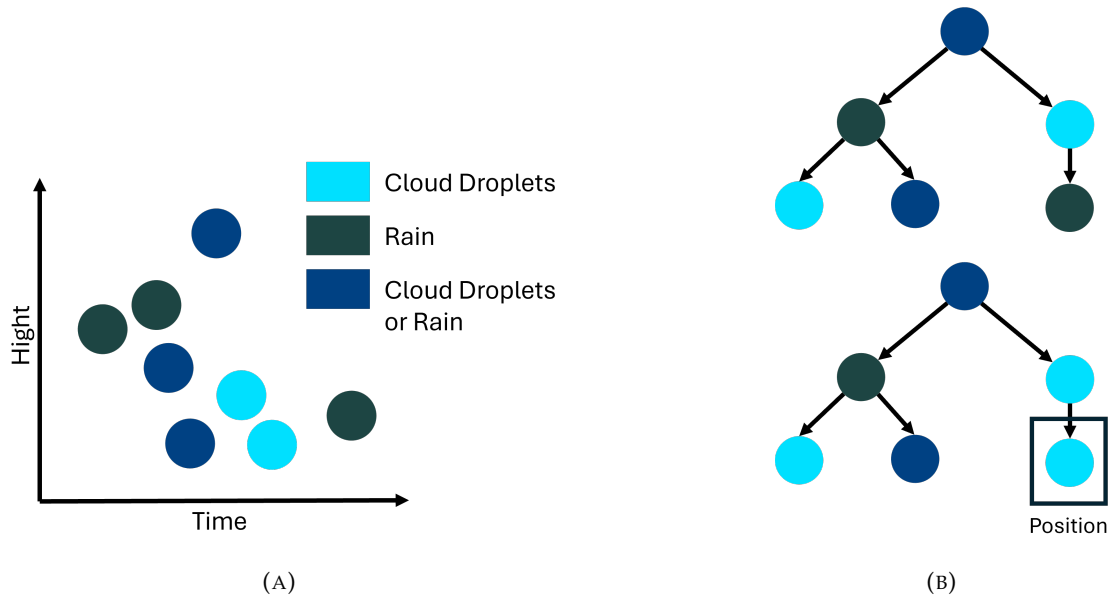


FIGURE 1.9: Overview of the research contributions on human perspectives on the use of color in data visualization for qualitative data.

1.3.2 Research Questions

To pursue this aim, this work answers the following research questions:

- **RQ2.1:** What strategies should be used to create multi-label class color schemes with strong semantic associations?
- **RQ2.2** Which factors influence the perception of differences across categorically colored node-link diagrams?

Understanding these strategies and influencing factors can enhance the discriminability of qualitative node colors and therefore the interpretability of the data presented. Hence, the answers to these research questions represent important advances toward the effective color representation of data attribute values in nodes and, consequently, toward effective color designs for node-link diagrams.

1.3.3 Contributions

More specifically, these papers make the following contributions:

- An algorithm for creating color schemes for multi-label data that aims not only to ensure the discriminability of colors but also to find color representations that correspond to users' semantic color associations (see [Figure 1.9a](#)).
- A collection of surrounding factors in a graph that positively and negatively influence the recognition of color differences when comparing those graphs. Additionally, an assessment of which of these factors differ from the recognition of differences in graphs without colors (see [Figure 1.9b](#)).

All findings are domain-agnostic and are applicable to various contexts. They are intended to provide a basis for future research that is urgently required in the domain of color perception in graph research.

1.4 Outline of the Thesis Structure

The remainder of this thesis is organized as follows (see [Figure 1.10](#)):

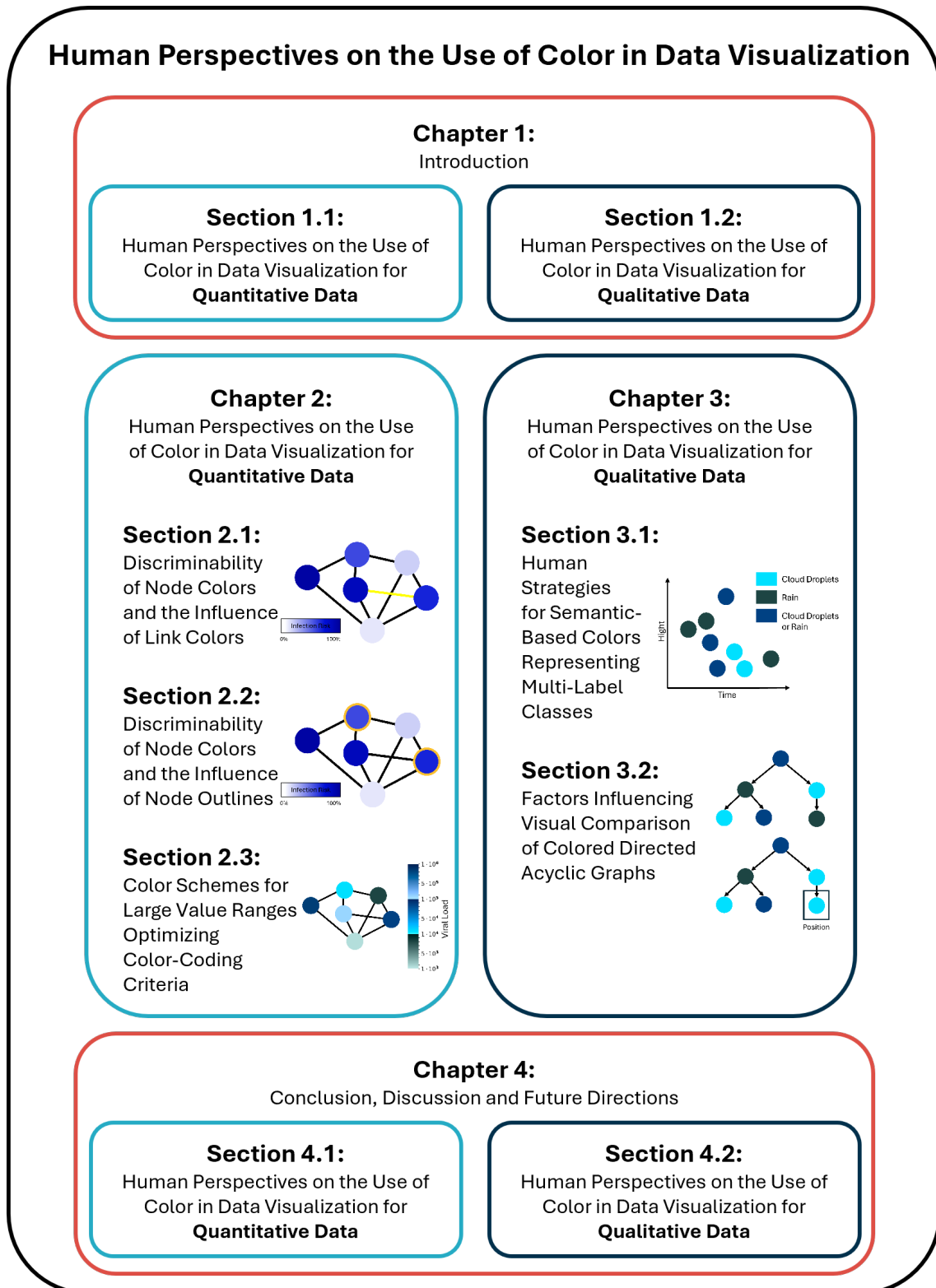


FIGURE 1.10: Outline of the thesis structure.

Chapter 2 contains results of the research on factors influencing the discriminability of node colors representing quantitative data attribute values. The three papers presented investigate the influence of different link hues and topologies ([Section 2.1](#)), the influence of different node outline designs ([Section 2.2](#)), and finally the third paper investigates various color schemes to represent large value ranges attribute values in node colors under consideration of visual deficiencies ([Section 2.3](#)).

Chapter 3 contains results of the research on factors influencing the discriminability of node and point colors representing qualitative data attribute values. The two papers presented investigate strategies for finding semantically resonant color schemes for multi-label data ([Section 3.1](#)) and factors influencing the discriminability of node colors across node-link diagrams ([Section 3.2](#)).

In **Chapter 4** the findings of both chapters regarding each data type being visualized are summarized and discussed. Furthermore, directions for future work are suggested.

Chapter 2

Human Perspectives on the Use of Color in Data Visualization for Quantitative Data

2.1 Discriminability of Node Colors in Node-Link Diagrams: Link Hues

2.1.1 Introduction

The following paper examines the effects of edge colors and topologies on the discriminability of node colors. It presents a series of studies that systematically vary edge colors and topologies while comparing nodes with different color differences, which study participants must identify. In doing so, the study determines so-called just noticeable differences, which are intended to illustrate the influence of the edges on the discriminability of node colors.

2.1.2 Just Noticeable Differences

Just Noticeable Differences (JND) refer to the smallest difference between two stimuli that can be reliably perceived by a viewer [49]. The JND is not a fixed difference but is statistically estimated from human responses using a function, which relates stimulus differences to the accuracy of detection [83]. Depending on the experimental setup, the JND is defined as the difference corresponding to a predetermined criterion. In visualization research, the criterion often used is the difference at which 50% of viewers can detect the displayed differences. To estimate this difference, viewers are presented with pairs of stimuli that differ by varying amounts and are asked to indicate whether the stimuli differ. The resulting function delivers a quantitative value of perceptual sensitivity [83,99].

2.1.3 Discriminability of Node Colors in Node-Link Diagrams and the Influence of Link Colors

This paper [114] was published in *Visual Informatics*, Volume 10, Issue 2, in June 2026:

Laura Pelchmann, Rita Borgo, Margit Pohl, Tatiana von Landesberger, Discriminability of node colors in node-link diagrams and the influence of link colors, *Visual Informatics*, Volume 10, Issue 2, 2026, 100288, ISSN 2468-502X, doi: [10.1016/j.visinf.2025.100288](https://doi.org/10.1016/j.visinf.2025.100288)

I am the primary author of this publication. In this role, I was responsible for the design, implementation, data collection and analysis, as well as the writing and publication of the work. The individual contributions of myself and my co-authors to this publication are outlined below according to the Contributor Roles Taxonomy (CRediT):

Laura Pelchmann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft.

Rita Borgo: Rita has made important contributions through her expertise in study design, implementation, and statistical analysis. Moreover, she was always available to provide feedback and answer questions throughout every aspect of the research process. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Margit Pohl: Margit has given important insights into psychological studies and their design, pertinent previous literature, and the appropriate variation of stimuli. Additionally, she consistently provided feedback and addressed questions about all aspects of the research process. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Tatiana von Landesberger: Tatiana has extensive expertise in designing graph comparison studies and has been instrumental in supporting me throughout the research project, offering crucial feedback and constantly aiding in decision-making processes. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Discriminability of Node Colors in Node-Link Diagrams and the Influence of Link Colors

LAURA PELCHMANN¹, RITA BORGO², MARGIT POHL³, TATIANA VON LANDESBERGER¹

¹*University of Cologne*

²*King's College London*

³*TU Wien*

Abstract:

This paper examines how link colors influence the discriminability of quantitative node color encoding in node-link diagrams. Node colors in node-link diagrams can represent quantitative attributes, such as value, frequency, or intensity. Although links themselves do not necessarily represent values, the relational aspect must be displayed with a color. This color has the potential to influence the discriminability of the node colors. Extensive research has been conducted on quantitative encoding for scatterplots; however, there is currently no research on which color combinations between nodes and links should be employed and how the quantitative values in nodes should be encoded to ensure their discriminability. This paper addresses two cases: 1. Comparing two highlighted nodes while all other nodes are gray, and 2. Comparing two nodes in a fully colored graph. We conducted two studies, in which participants were required to identify color differences between two nodes in a node-link diagram with varying link properties. In the first study, conducted online, two nodes were colored, and the surrounding link topologies and link hues were varied. The node hues were also varied, while the quantitative values were encoded through saturation. The most discriminable node hues were used in the second study, conducted in a laboratory setting, where all nodes encoded quantitative data. Our findings indicate that the use of complementary-colored links has the effect of enhancing the discriminability of node colors, irrespective of the underlying topology. In contrast, links with a hue that is similar to the node hues leads to a reduction in discriminability. The findings of both studies were found to be consistent. In sum, our results indicate guidelines for the use of colors to enhance the discriminability of nodes in small graphs with quantitative node encoding. We recommend using shades of blue rather than yellow to quantitatively encode nodes. To highlight the connections between nodes, pair these colors with complementary colors for the links, rather than coloring the links the same color as the nodes. Alternatively, use neutral colors, such as gray, to support the discriminability of node colors.

1. Introduction

Color encoding represents an effective method for the visual representation of quantitative data in node-link diagrams [26,55,95,127,147]. For example, in contact networks, individuals can be represented as nodes, with the colors indicating people's infection risks [94]. This may be applied to two highlighted nodes, which could represent infection transmission risk,

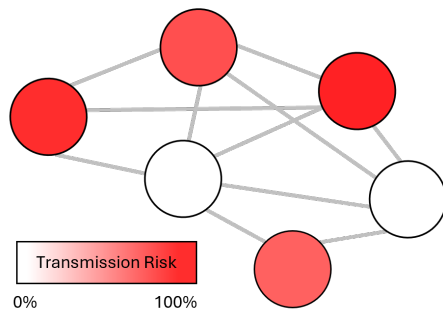


FIGURE 1: Shows an example of a contact network that represents disease transmission risks.

or alternatively, it could be used to color all nodes, indicating the overall infection risk for surveillance purposes (see Figure 1). An accurate representation of the numerical values of these nodes requires that the colors be discriminable, i.e., that the perception of color differences corresponds to the perception of value differences between pairs of nodes. Therefore, node coloring and discriminability are important for graph drawing exploration.

Node-link diagrams depict relationships through links, which can affect the perception of node colors by adding a relational and proximity component. The unique features distinguish these diagrams from non-relational visualizations, such as scatterplots. It also motivates a closer examination of their perceptual dynamics, since they need to be represented by a color, as well as the nodes, even though they might not necessarily represent values. Figure 2 shows a motivating example of this effect. It combines the different node hues with the edge colors and topologies used in the experiments. Our work wants to investigate which representation of link color and node color is best suited to better discriminate node colors.

The visual detection of color differences is a topic that has been widely studied [29,150,152,170]. While previous studies have investigated perceptual color differences for various chart types [102,152,159], they have mostly neglected to consider node-link diagrams. Color perception of nodes in node-link diagrams can be considered analogous to the perception of points in scatterplots. Thus, the most related studies have examined the discriminability of color differences between differently sized highlighted points in scatterplots by Szafir [152]. However, to the best of our knowledge, it is still unknown how the links placed between nodes influence the differentiation of node colors.

The following questions guided our research:

- How do links between nodes impact node discriminability in node-link diagrams? How do different combinations of link topologies influence node color discriminability? This shows whether the color guidelines from scatterplots are applicable to graph drawing.
- How do link and node color choices (i.e., hue) impact the node color discriminability in node-link diagrams? This guides the choice of appropriate combination of link and node hues for graph drawing.

We conducted two experiments in which participants identified color differences between nodes in small node-link diagrams. First, we conducted a study with two colored nodes, followed by a study with a graph containing nodes of various shades of blue. Our goal was to gain initial insights into this new area of research. This study focuses on the

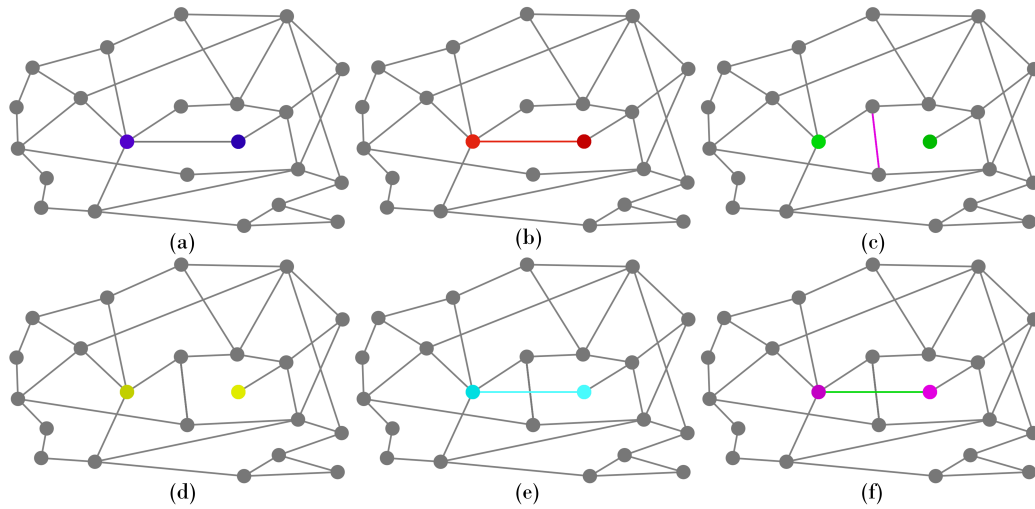


FIGURE 2: These examples from our first study demonstrate that the detectability of color differences between nodes is affected by link colors (gray: (a)+(d), same color: (b)+(e), complementary colored: (c)+(f)), link topologies (connecting: (a)+(b), separating: (c)+(d), cross: (e)+(f)), as well as the node colors (blue: (a), red: (b), green: (c), yellow: (d), turquoise: (e), pink: (f)). The two nodes at the center of each node-link diagram have the same difference in color values. However, our studies show that people do not detect them equally well.

basic case of encoding quantitative values in nodes of one size with no encoding in links. We limited the number of distractors in terms of the coloring and topology of the inserted links, as well as the width of the links. This lays the groundwork for future extensions involving other encodings, such as categorical node colors, link colors, node shapes, and sizes.

Our contribution is the finding that link coloring significantly affects the discriminability of node colors. When the links were colored in a complementary hue to the nodes, differences were perceived more clearly than when the link color was within the same hue as the nodes. In addition, the colors of the nodes were more discriminable when displayed in blue rather than yellow. Our results indicate guidelines for drawing colored node-link diagrams.

The following section presents a review of related research, followed by a description of the studies. Following the analysis and results, we present the findings and provide guidelines for the use of color in node-link diagrams. We conclude with implications for future research.

2. Related Work

We provide an overview of research on graph perception and the perception of color in visualizations.

2.1 Graph Perception – Links and Colors

Research on color differences is still under-investigated in graph drawing. In particular, the perception of color in node-link diagrams is still a rare topic [55]. Burch et al. [27] state the aspect of color in node-link diagrams as a challenge.

The versatility of graph data and the visual characteristics of node-link diagrams play a key role in their perception [101]. They also have influence on human understanding of the data and complexity [122, 149, 173]. Saket et al. evaluated the ability to represent relationships in node diagrams, node-link diagrams, and node-link-group diagrams [134].

They found that the addition of links or link and group representations had no negative effect on the performance on node-based tasks. For example, the effect of rounded links [68] or the effect of link crossings on understanding the data [69]. Therefore, the topology of added links between the nodes plays an important role in our experiments. Reimann et al. [128] investigate the influence of colored links in clustered node-link diagrams. They investigate how colored links can be used to better distinguish clusters of people. The color discriminability of the nodes is not studied. This kind of research influenced our research questions.

2.2 Perception of Color in Visualization

The effective use of color in visualizations has long been a central area of research [168]. Various types of research have been conducted, including studies on how colors can represent data, the unique understanding of these colors, and their optimal assignment [36, 91, 98]. An important factor contributing to the uniqueness of a color is accurately mapping different numerical values to colors [26, 95, 127]. In order for these value-color combinations to be meaningful, color scales are required that place discriminable colors in a logical context, depending on the data type and number of data values. Several tools can help us select colors to create a color scheme that depends on the data type and number of data values [54, 57, 106, 154]. But, these tools can only assist the user in choosing colors and do not guarantee a discriminable representation of values or attributes. Discriminability between the assigned colors plays an important role [152, 167].

Studies of maximum discriminability between groups of points with the same attributes are usually based on a predefined set of colors with known color differences that are to be optimally arranged. This does not necessarily allow conclusions to be drawn about the discriminability of colors in general, as other aspects, such as the color of the background of a chart [141] or different colored areas in the proximity may influence the discriminability [102, 170, 172]. The theories of contrasting colors can also be applied here, as described, for example, in [33, 80, 129, 169]. Thus, we study the factor of proximity of nodes and links as well as link color on node color discriminability.

Nodes in node-link diagrams correspond to points in scatterplots connected by links. An example for this correspondence is GraphDice [9]. Nodes in node-link diagrams are placed according to their attributes as points in scatterplots. They are connected by links to create a node-link diagram. This results in the idea to use research findings on scatterplots as a baseline for our experiments. The optimal use of categorical color schemes has been the subject of several research projects, including investigations into the best possible distribution of predefined color sets [167], the optimal highlighting of individual aspects in multi-class scatterplots [96] or the extent to which the number of colored categories can be discriminated by humans [158]. In contrast, we focus on quantitative value visualization as in studies by Szafrir [152]. She presents a set of functions that describe the JNDs for bar charts, line graphs, and scatterplots as a function of the size of the bars, lines, and points. For highlighted dots in scatterplots, participants decided if they were of the same color or not. Our own research questions are therefore aligned with those posed by Szafrir [152]. As established by Szafrir [152], most experiments attempting to gain insights are performed under controlled laboratory conditions with perfect visual conditions. Therefore, the significance of the JNDs found by the staircase method [37], for example, is doubtful for applied visualizations. Our research methodology is inspired by Szafrir's methodology [152].

3. Research Questions

The main research goal is to evaluate if and when links affect node color discriminability in node-link diagrams.

We begin with an examination of the influences of link topologies to show the specific differences and similarities compared to scatterplots. We then conduct an investigation of link hues with the goal of identifying the most significant hues in the context of the different topologies. The resulting research questions **RQ** and hypotheses **H** are:

RQ1: *Does the presence of links in node-link diagrams affect how discriminable the node colors are compared to scatterplots?*

H1: Links in node-link diagrams, in contrast to scatterplots, make discrimination of node colors more difficult, because links affect the viewer's attention.

RQ2: *Do different link topologies between two nodes have different effects on the discriminability of the node colors?*

H2: The perception of color differences between two colored nodes is easier when there are no links between them. The presence of link crossings between the two nodes makes discrimination more difficult.

RQ3: *Do different link hues between two nodes have different effects on the discriminability of the node colors?*

H3: A gray link between two nodes improves discrimination more effectively than a link within the nodes' hue range. However, a link with a complementary hue to both nodes may negatively affect the discriminability of the node colors due to the high contrast between the link and the nodes on both ends of the link and possible effects on the color perception.

RQ4: *Do different link hues between two nodes have different effects on the discriminability of the node colors when they are applied to different link topologies?*

H4: The influence of different links hues on separating links is less pronounced, and the resulting effects are accordingly less variable than in cases where more links are present between the two nodes.

RQ5: *Do different hues of two nodes affect their discriminability when links of different topologies and hues are between them?*

H5: Due to the color properties and perceptual capabilities of the human eye [119], color differences between green-colored nodes are easier to perceive than those between red-colored nodes, which are easier to perceive than those between blue-colored nodes.

4. Study Design

We conducted two studies with comparable stimuli and tasks. First, we investigated the discriminability of two colored nodes in a graph with only two colored nodes other than gray (calling them highlighted nodes). Then, we investigated the discriminability of a graph with all nodes being colored. The node highlighting setup corresponds to the point highlighting in Szafir's experiment. We also use it for comparison purposes.

Task In both experiments, participants were shown a picture of a node-link diagram and asked to decide if two colored nodes in the center were the same color or different colors (example pictures shown in Figure 3b).

Variables The recognition of color differences is the dependent variable. In our studies, we varied the link topology, link hues and node hues (see Figures 3-5), which are our independent variables.

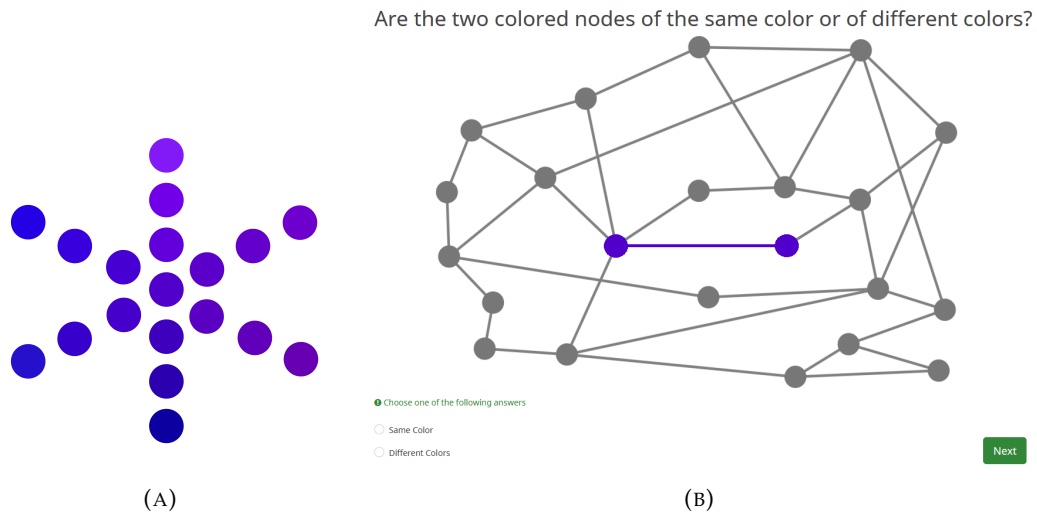


FIGURE 3: Figure (a) shows the 19 node colors for the blue hue resulting from changes along every CIE $L^*a^*b^*$ axis in both experiments and Figure (b) shows an example screenshot of Study 1.

We explain the similarities/differences between the two studies and then go into the details of each study.

4.1 Stimuli Design

Node-Link Diagram We created a node-link diagram that allows for comparing results to Szafir's scatterplot experiments [152]. This means that the number of nodes and the node positions, as well as their coloring are comparable to their setup. The 22 nodes had a size of 25 pixels. We then added 6 pixel wide straight links to create a connected graph as usually used in graph perception experiments [120, 121]. We added links so that the graph layout would resemble the results of a force-directed layout algorithm. The layout should adhere to aesthetic criteria such as minimizing link crossings, large link angles, similar link lengths, and similar node distances as much as possible (see Appendix 3).

For experimental purposes, we specifically varied the links between two selected nodes in the graph (i.e., the highlighted nodes). In order to vary the links that directly affect the connection between the two colored nodes, there were no links between them in the middle of the node-link diagram, neither connecting the nodes nor placed between them (see Figures 4a and 4e). The resulting graphs were rendered on a white background with a size of 1205 x 805 pixels using D3 [18].

We colored nodes and links as described below. The default node and link color for unhighlighted objects was gray.

Color Differences Regardless of the color arrangement, measurability must be ensured in order to study color differences and their perceptibility. The CIE $L^*a^*b^*$ color space was developed to provide this measurability [100]. It is a three-dimensional color space (L , a , b) designed to represent perceptually accurate distances between colors and to make them measurable. The a^*b^* plane employs opposite color theory. Green and red are opposite each other on the a^* -axis. The b^* -axis runs between blue and yellow. The L^* -axis describes the color's luminance. The mathematical formula for the distance between three-dimensional points can be used to measure these distances [71].

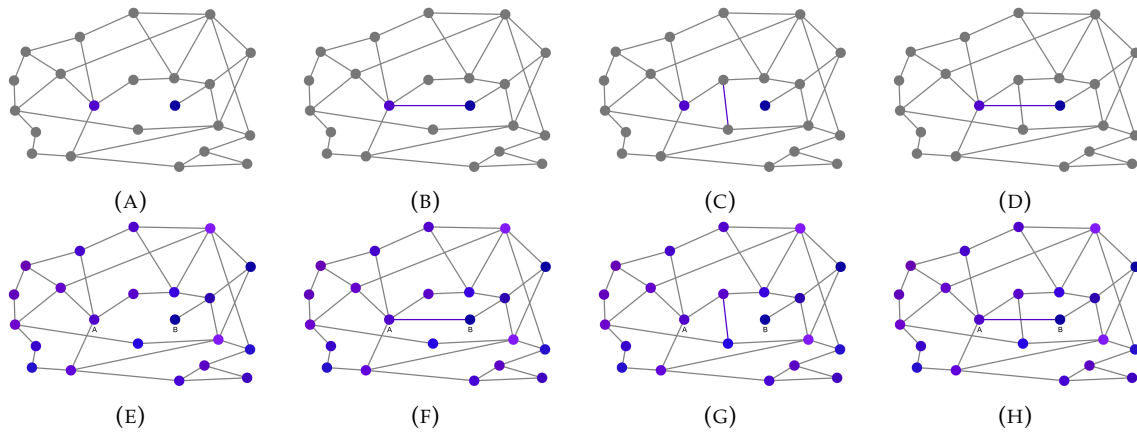


FIGURE 4: Figures (a)-(d) show the variations of link topologies with links colored in blue examined in Study 1. Figures (e)-(h) show the same variations examined in Study 2.

For the purpose of reaching our research goal, the colors were slightly modified along all three axes of the CIE $L^*a^*b^*$ color space, starting from the selected colors. Here we deviate slightly from Szafir's setup [152]. Instead of testing differences of $\Delta E = 6$ and $\Delta E = 12$ in both directions of each CIE $L^*a^*b^*$ axis, we used differences of $\Delta E = 5$, $\Delta E = 10$ and $\Delta E = 15$ to show a wider range of differences and thus perform more accurate calculations. Using this variation method, we obtained 19 different color stimuli for each hue to compare, constantly showing one node in the base color and varying the other in one of the 19 colors. An example for the resulting color variations can be seen in Figure 3a.

Over the years, several new formulas have been developed [131, 144]. To maintain comparability with state-of-the-art work on color differences [144], our analysis uses two formulas: ΔE and CIEDE2000(ΔE_{00}).

Link Topologies To test the effect of different link topologies on the discriminability between two colored nodes in a node-link diagram (RQ2) we varied four topologies between the two colored nodes:

- No links between the two nodes¹ ("None", see Figure 4a).
- One link connecting both nodes ("Connecting", see Figure 4b).
- One link separating the space between the nodes ("Separating", see Figure 4c).
- Two links crossing each other. One connecting both nodes, one separating the space between them ("Cross", see Figure 4d).

These topologies represent, in a simplified way, possible cases of links or lack thereof between two nodes. Two nodes may be unconnected, in which case there are either no links between them, or links connecting other nodes. If two nodes are connected, this connection exists either with or without link crossings.

Link Hues As with topologies, the link hues represent the different color schemes used in node-link diagrams. They are either near neutral or harmonic in hue or contrast [33].

¹The surrounding links remain unaltered.

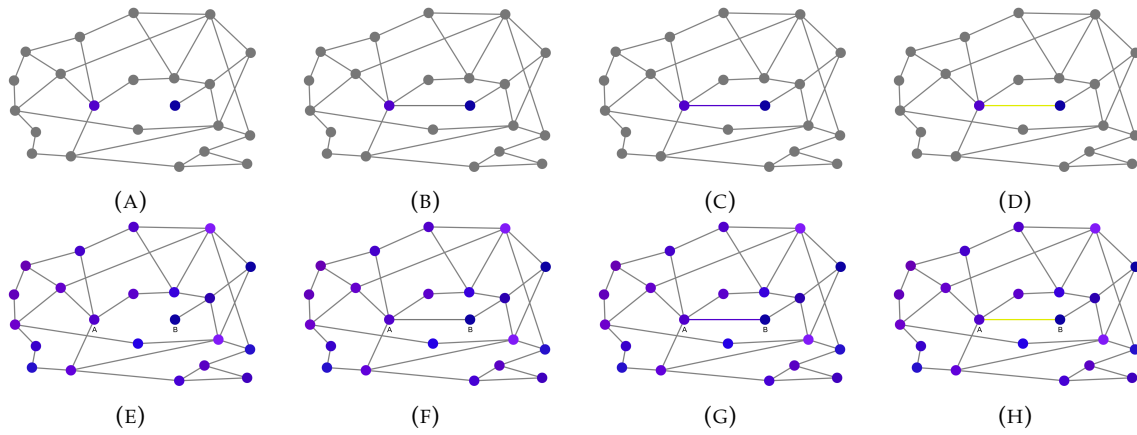


FIGURE 5: Figures (a)-(d) show the variations of link hues with consistent link topologies examined in Study 1. Figures (e)-(h) show the same variations examined in Study 2.

To test the effect of these different link hues between the two colored nodes (**RQ3**) we chose:

- No inserted link. ("None", see Figure 5a).
- Links colored in gray ("Gray", see Figure 5b).
- Links with same hue as at least one of the nodes ("Same", see Figure 5c).
- Link hue complementary to the nodes ("Complementary", see Figure 5d).

Gray links represent the default natural, near neutral link color choice for the other links in our node-link diagrams. Same and complementary hues evaluate the effect of object color in proximity to target objects (see Section 2).

Node Hues Study 1 varied six different node hues (see Figure 2 and **RQ5**). We chose the three primary colors of the additive color space (red, blue, and green) for these colors. Since we also wanted to test the influence of complementary colored links, we included turquoise, yellow, and pink as node hues, covering a span of [25, 90] on the L^* -axis, [-75, 95] on the a^* -axis, and [-90, 90] on the b^* -axis. These ranges provide a solid foundation for expanding to more colors in the entire color space in the future. The six hues were divided into six parallel experiments. Within each experiment, the hue just slightly altered. As described in 4.1, we chose three alterations along both directions of each axis of the CIE $L^*a^*b^*$ color space, resulting in 19 different hues per experiment (see Figure 3a, the exact color values are presented in Appendix 2). With the 6x19 different hues we chose, we wanted to get a first impression of different hues that can hopefully be extended to the whole color space in future work similar to [152]. The best hue result was used in the follow-up study.

4.2 Study 1: Color Discriminability of Two Highlighted Nodes

Study 1 assessed a node-link diagram with two highlighted nodes whose colors were to be compared. We varied link topology, link hues and node hues (see above). It resulted in a within-subject study (node hue groups), in which each participant had 190 trials (i.e., 3 topologies x (3 hues + no links) x 19 color stimuli).

Procedure Similar to Szafir, the survey was conducted online. We used LimeSurvey [52]. To view the diagrams at the correct size, we asked participants to use a computer screen and avoid smaller displays, such as smartphones. Color vision testing with Ishihara plates was followed by a short training session with explanations and two trial questions. One of the explanation questions showed the node-link diagram used in the experiment, with two nodes of the same color and one with two nodes of different colors. This was followed by a warning that the color differences shown were not necessarily representative and that the study pictures should be judged with great caution.

All 190 images were presented in a randomized order to minimize learning effects. A break was offered after 95 pictures to reduce the fatigue effect. Finally, participants provided information about their mother tongue, experience with graphs, age, and gender identity. Meaning and understanding of colors may differ across cultures [2], which may have an impact on the detection of differences between colors. The question about the mother tongue was asked to determine possible differences in the participants' cultural backgrounds.

Participants could exit the survey at any time. The average completion time per participant was 24 minutes. Participants spent an average of six to eight seconds per question.

Participants A total of 66 participants (30 male, 34 female, and 2 prefer not to say) finished all 201 questions². Participant's ages were distributed between 18 and 69, with the majority (80.3%) being between 20 and 40 years old. To avoid possible influence of color blindness, we filtered out participants (2 persons) failing the Ishihara test [73]. Participants were recruited through calls for participation at universities among students and in research groups. Most participants (87.8%) had some experience with graphs.

There was no bias on graph expertise between study groups due to random assignment. The χ^2 -test showed no statistical evidence that hues and expertise correlate ($\chi^2 = 13.452$, p -value = 0.857).

4.3 Study 2: Color Discriminability in a fully colored graph

We studied the general case where all nodes of the graph were colored, rather than just two being highlighted. Given the potential for screen settings to influence results in online studies, Study 2 was conducted in a laboratory setting with controlled conditions.

Stimuli The stimuli chosen were the same as in the previous study, see Section 4.1. Study 1 revealed that blue nodes were the most discriminable among the tested colors. Therefore, to maintain consistency with Study 1, we decided to color all nodes blue and leave the possible link hues as gray, blue, and yellow in Study 2. Unlike Study 1, all nodes in the node-link diagram were colored. The color differences between nodes and the hues and topologies of the links were identical to those shown in Study 1.

Procedure The study was conducted in a laboratory environment. The study was presented on a 27-inch screen calibrated for true-color display using OLED technology. The room was darkened during the study and the participants were left alone after the introduction. Participants then again had to decide in each of the 190 trials ((3 topologies $\times$ 3 hues + no links) $\times$ 19 color stimuli) whether the two colored nodes displayed were of the same or different colors. The average time was 22 minutes. All 190 images were again presented in a randomized order to reduce the potential for learning effects. Participants

²Before starting the online study, we did a pilot study with six participants.

spent between six and eight seconds per question on average. Finally, we asked the participants to rate all 10 different link variations with a Likert Scale from "Very difficult" and "Difficult" over "Neutral" to "Easy" and "Very easy". We also asked them to explain their strategies and what would have decreased the difficulty.

Participants The 15 participants (7 male, 8 female) were recruited from university students and research groups, none of whom participated in the first experiment. Most participants (93%) had at least some experience with graphs. The number of participants is state of the art for this type of study [10,48,61].

5. Analysis Methodology

The method of analysis for comparability follows the calculations in [152]:

1. **Testing for normal distribution:** For each of the six experiments, both the Kolmogorov-Smirnov test (sign. < .001) and the Shapiro-Wilk test (sign. < .001) rejected the null hypothesis that the data were normally distributed.
2. **Testing for independence:** For each of our research questions, the Kruskal-Wallis test was used to test the independence of the factors under investigation.
3. **Discriminability rate:** At the beginning of the calculations, the discriminability rate was calculated for each of the 190 questions in the experiment. This rate is determined by the following formula: $DRate = \frac{\text{amount of detected differences}(\Delta E)}{\text{amount of total trials}(\Delta E)}$.
4. **ND(50%):** With the resulting discriminability rate per color difference displayed, we calculated a linear regression. Following Szafir's calculations [152], we forced the regression line through the origin, because the amount of people detecting differences between two identical colors should be zero. We used the formula: $m_{axis} = \frac{\Delta E * DRate(\Delta E)}{DRate^2(\Delta E)}$.
The slope of the linear regression line was then used to determine the estimator for noticeable differences per axis: $ND_{axis}(p) = \frac{p}{m_{axis}}$.
Here p is the desired factor of the discriminability rate, i.e., the proportion of people who can tell the difference between two colors. $p = 0.5$ is often used for the investigation of just noticeable differences [152]. Thus, we report the ND(50%) value.

The same calculation was performed for both the ΔE and the ΔE_{00} distances between the colors.

6. Results of Study 1: Highlighted Nodes

Study 1 analyzes the effects of links in a setup similar to Szafir's – with highlighted objects. First, we analyze the general influence of links comparing node-link diagrams to scatterplots. And then we analyze in more detail the effect of link topologies, the influences of link colors as well as link colors per link topology and different node colors.

The result section shows the ND(50%) values calculated as described above, as well as the results of statistical tests. Further statistical tests and their results are provided in the Annex.

6.1 General Link Influence

The present study analyzes the general influence of the links, i.e., the discriminability of node colors in node-link diagrams versus the discriminability of colors of points in scatterplots. To obtain direct comparison values with Szafir's calculations, we restrict

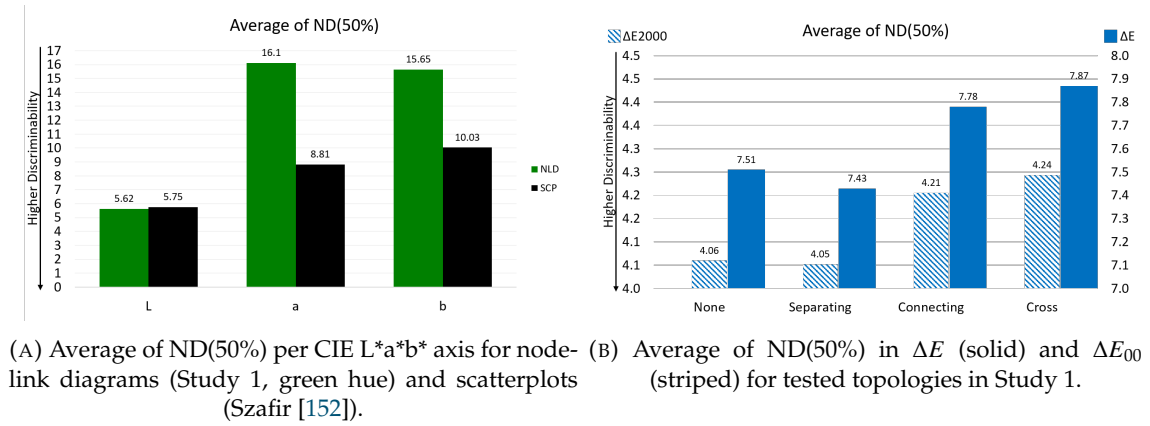


FIGURE 6

our analysis to the results of our green hue study. Figure 6a shows a comparison is presented between the results obtained in this study and those reported by Szafir [152]. To this end, the required color differences for 50% of the participants to detect them are divided into values for each of the CIE L*a*b* axes. It shows a large increase in the ND(50%) values for the a and b axes, as well as for the total. This finding suggests that participants experienced greater difficulty detecting color differences in the node-link diagrams compared to the scatterplots when the color was modified along the a and b axes in the CIE L*a*b* color space. However, no significant differences were observed for changes in luminance (along the L axis). Subsequent results: The discriminability rate, the resulting ND(50%) and the perceptible differences per axis and per link variant for our two chosen distance measures are shown in 12. Overall, the hypothesis **H1** was confirmed.

In order to be able to discuss the individual characteristics of our research questions in more detail, we calculated and use the average color changes independently of the axes.

6.2 Link Topology

The independence tests between tested link topologies showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) with result = .007 < .050 rejected the null hypothesis of dependent categories.

Figure 6b shows that the color discriminability varies significantly depending on the link topology. The resulting average color difference for a separating link is only $\Delta E = 7.43$ ($\Delta E_{00} = 4.05$). An average difference of $\Delta E = 7.78$ ($\Delta E_{00} = 4.21$) for connecting links and $\Delta E = 7.87$ ($\Delta E_{00} = 4.24$) for crossing links. This means that larger color differences are required for neighboring nodes to be discriminable by the observer. Especially for nodes with multiple links between them, it is important to ensure discriminability by choosing colors with large differences. This confirms **H2**.

The response time did not vary significantly across the different topologies. Kruskal-Wallis, Asymptotic Significance (2- sided test) = .392 > .050, retaining the null hypothesis of dependent categories.

6.3 Link Hues

The independence tests between tested link hues showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = .033 < .050, rejecting the null hypothesis of dependent categories.

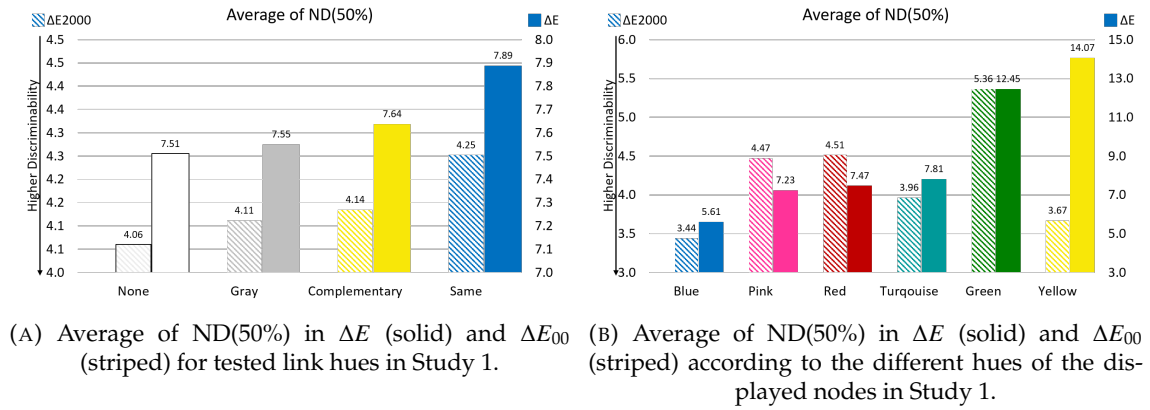
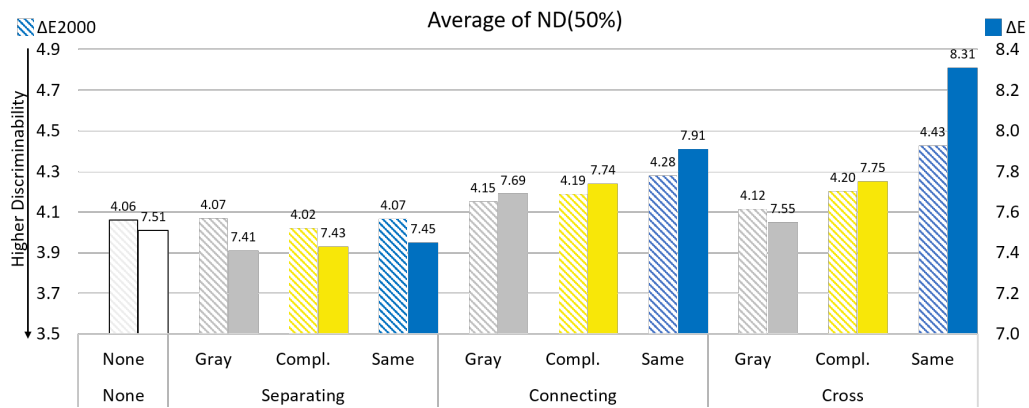


FIGURE 7

FIGURE 8: Average of ND(50%) in ΔE (solid) and ΔE_{00} (striped) for tested link hues and link topologies in Study 1.

The response time did not vary significantly across the different link hues. (Kruskal-Wallis, Asymptotic Significance (2- sided test) = .106 > .050, retaining the null hypothesis of dependent categories.)

Figure 7a shows that color discriminability (ND(50%)) varies significantly with link hues. The average color difference for a gray link is only $\Delta E = 7.55$ ($\Delta E_{00} = 4.11$). At the same time, for a link in close proximity colored in the same hue as the two colored nodes, the calculated average distance needed is $\Delta E = 7.89$ ($\Delta E_{00} = 4.25$). The first aspect of **H3** could thereby be confirmed, while an average distance of $\Delta E = 7.64$ ($\Delta E_{00} = 4.14$) for complementary colored links rejects the second aspect of **H3**. This means that larger color differences are required for discriminability. Moreover it is beneficial to color nodes and links in different or even complementary hues.

6.4 Link Hues per Topology

The independence tests between each tested link topology-hue-combination between the two colored nodes, showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = .010 < .050, rejected the null hypothesis of dependent categories.

Figure 8 shows that the color discriminability varies significantly. The resulting average color difference for a separating gray link is only $\Delta E = 7.41$ ($\Delta E_{00} = 4.07$). At the same time, for a link crossing, with the separating link colored in gray and the connecting link colored in the same hue as the two nodes, the calculated average distance is $\Delta E = 8.31$

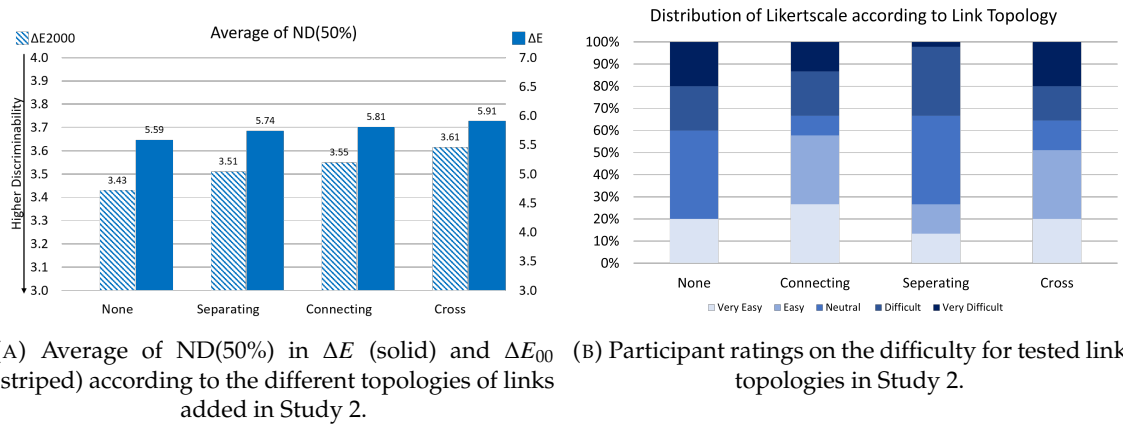


FIGURE 9

($\Delta E_{00} = 4.43$). Furthermore, the three tested hues show minimal differences in their effects on the links separating the two nodes. However, there are notable differences in the effect of hue on the links connecting the two nodes, confirming **H4**.

6.5 Node Hues

The independence tests between node hues, showed a significant effect (Kruskal-Wallis, Asymptotic Significance (2- sided test) = .001 < .050, rejected the null hypothesis of dependent categories.)

Figure 7b shows significant difference between hues. The resulting average color difference for two differently colored blue nodes is $\Delta E = 5.61$ ($\Delta E_{00} = 3.44$). At the same time, for two colored yellow nodes, the calculated average ΔE distance needed is $\Delta E = 14.07$. For the ΔE_{00} analysis the green hue showed to be the most difficult to spot differences in resulting in $\Delta E_{00} = 5.36$ for nodes colored in a green hue. These results we could not confirm **H5**. We also recommend using blue rather than yellow to represent important quantitative node attributes in node-link diagrams.

7. Results of Study 2: Fully Colored Graph

The evaluation for Study 2 was divided into two parts. First, the quantitative analysis that was analogously done for Study 1 was performed, and second, we evaluated the difficulty ratings given during the interview.

The participants in the second study detected more differences than in the online study. Looking at the average scores of only those participants who completed the online study with the blue colored nodes, the results are comparable. This suggests that the combination of the use of blue colored nodes and a better screen is responsible for the increase in detectable differences in Study 2. However, the results themselves, i.e., the differences between each topology and the colors of the inserted links, are very similar.

7.1 Link Topology

The independence tests between different link topologies between the two colored nodes showed no significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = .190 > .050 was not able to reject the null hypothesis of dependent categories. Figure 9a shows the ND(50%) results.

The results show that the color discriminability varies depending on the link topology. The average color difference for no link is $\Delta E = 5.59$ ($\Delta E_{00} = 3.43$). At the same time, for a link crossing, the value is $\Delta E = 5.91$ ($\Delta E_{00} = 3.61$). It shows that larger color

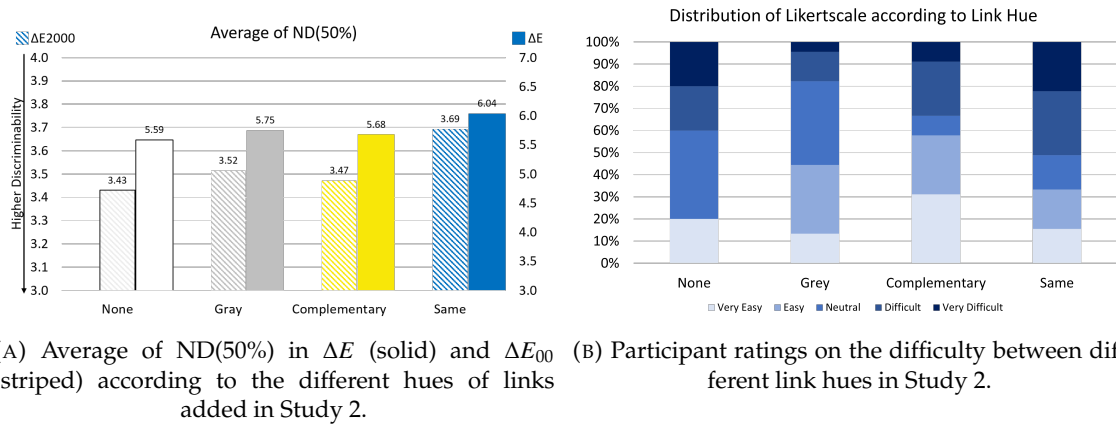


FIGURE 10

differences are required for neighboring nodes to be more easily recognized by the observer. Especially for nodes with multiple links between them, it is important to ensure discriminability by choosing colors with large differences. This confirms **H2**.

The participant's subjective difficulty ratings showed that no link between the two nodes *A* and *B* increased difficulty of color discrimination. The answer "Easy" was not chosen at all and only few people chose "Very easy" (see Figure 9b). This is in contrast to their performance. Although they found it difficult, participants still detected many color differences when no link was between the two nodes.

7.2 Link Hues

The independence tests between each added link hue between the two colored nodes of the node-link diagram, showed a significant effect of the different hues. (Kruskal-Wallis, Asymptotic Significance (2- sided test) = .007 < .050, rejecting the null hypothesis of dependent categories.)

The test for independence of response times as a function of link hues was the only one to produce a statistically significant result. (Kruskal-Wallis, Asymptotic Significance (2- sided test) = .008 < .050, rejecting the null hypothesis of dependent categories.)

Figure 10a shows that the color discriminability varies significantly with link hues. The resulting average color difference for no link between the two nodes is only $\Delta E = 5.59$ ($\Delta E_{00} = 3.43$) while it is $\Delta E = 6.04$ ($\Delta E_{00} = 3.69$) for links colored in the same hue as the nodes. An average distance of $\Delta E = 5.68$ ($\Delta E_{00} = 3.47$) for complementary colored links supports rejecting the second part of **H3**. This means that larger color differences are required for nodes similarly colored as the surrounding links in order to ensure discriminability. Moreover it is beneficial for the human perception of color differences to color nodes and links in different hues.

A statistically significant result was obtained for the different link hues, in contrast to link topologies. This suggests that the coloring has a slightly greater effect on the perception of color differences of colored nodes than the link topology between the nodes.

The subjective difficulty ratings showed that complementary colored links increased discriminability in contrast to links colored in a similar hue (see Figure 10b).

7.3 Link Hues per Topology

Figure 11 shows the values of the quantitative analysis for all link topology-hue combinations. The results show that the required color differences between the two nodes vary significantly depending on the link topology-hue-combination. The resulting average color difference for no link between the two nodes is $\Delta E = 5.59$ ($\Delta E_{00} = 3.43$). At

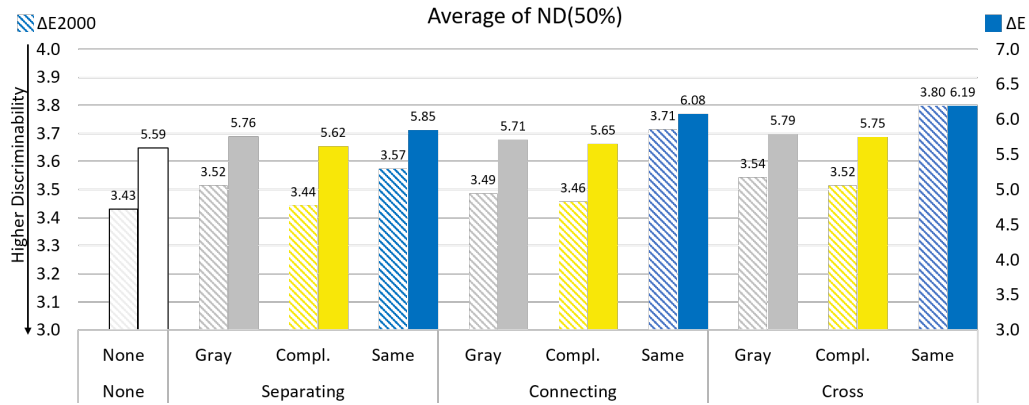


FIGURE 11: Average of ND(50%) in ΔE (solid) and ΔE_{00} (striped) according to the different hues of links added, subject to the different topologies of links added in Study 2.

the same time, for a link crossing, with the separating link colored in gray and the connecting link colored in the same hue as the two nodes, the calculated average distance is $\Delta E = 6.19$ ($\Delta E_{00} = 3.80$). The differences between the various hues are notably less pronounced for separating links in comparison to connecting links or crossing links – confirming **H4**.

8. Guidelines

The results of our experiments indicate the following guidelines for creating small node-link diagrams in which the colors of the nodes represent quantitative values as compatible with our studies:

Large Color Differences within Color Schemes: The presence of connecting links between nodes in node-link graphs makes it more difficult to differentiate between node colors. Therefore, we recommend large color differences within the color scheme used. This means that value differences must be represented by higher color differences than in scatterplots. (see Sections 6.2 and 7.1).

Neutral or Complementary Colored Links: If node attributes are displayed by color in a node-link diagram, it is recommended that links between nodes be gray. Neutral colors are the least distracting from the colors of the nodes and are therefore the preferred choice over the other tested colors (see Sections 6.3 and 7.2). In case of the need to highlight the links of a node-link diagram we recommend using complementary hues for the links to the nodes. This allows us to draw more attention to the links while still ensuring that the node colors are discriminable.

Blue Nodes: For coloring nodes in node-link diagrams, we recommend using blue color schemes rather than yellow for white backgrounds (see Section 6.5). These blue nodes provide the highest discriminability between different nodes and the values they represent among our tested hues.

Change in Brightness: The comparison to Szafir's results [152] shows that differences between the values represented in nodes are most effectively represented by changes along the L-axis of the CIE $L^*a^*b^*$ space, similar to scatterplots. Variations in lightness and darkness were most effective in our participants' perception of differences.

9. Limitations and Future Work

We conducted two studies on node-link diagrams: one with highlighted nodes and one with fully colored nodes. Our experiments varied node hue, link hue, link topology to obtain in-depth insights into the matter.

As this is one of the first studies of node color discriminability in node-link diagrams, we focus on the basic case of encoding quantitative values in nodes, but not in links. This leads to preliminary implications and lays the groundwork for future extensions involving other encodings, such as categorical node colors, link colors, and node shapes and sizes, as well as their combinations [110, 160]. Investigations containing applied, graph-based tasks, such as cluster identification, path finding, etc. [85], would also be interesting.

The design was chosen to be comparable to Szafir's scatterplot study. We examined six node hues, each of which consists of 19 colors. We also examined four link topologies, each of which consists of links of one thickness, as well as four link hues. This is a subset of all possible distractors within a node-link diagram. It is also a subset of the entire color space, limited to the colors we selected to minimize visual complexity. This limits the generalizability of the results to variations within the larger color space. However, adding more colors might hinder participants' ability to comprehend the visualizations [123]. The combination of these two factors raises important questions for future studies.

It would also be interesting to see how the results of the study change when the two nodes being compared are farther apart than those examined in the graph. For example, compare other nodes within the graph. Both studies used stimuli with one node size and one link width. Similar to Szafir's experiments [152], it would be interesting to investigate the influence of node size and link width.

Our study methodology was chosen to achieve comparability with method and results by Szafir [152]. However, in the future, it would be interesting to extend our experiments and use other methods, such as the staircase method, to detect significant differences. These results could then be compared to ours.

Another interesting aspect would be to consider the cultural and scientific backgrounds of the participants, as previous research has shown that these backgrounds influence color perception [82, 140]. Although we recorded the participants' native languages, our sample was not diverse enough for this factor to be evaluated in our experiment.

The position of the nodes was guided by the scatterplot study and force-directed layouts. This is the standard layout with high quality of graph aesthetic criteria. It is suitable for general small graphs with low density as in our case. This case is also widely studied in perception studies for graph drawing [120, 121]. Further investigation of color discriminability for other stimuli, such as the influence of graph density, graph size, and graph structure with layout variations is future work.

Our studies display node-link diagrams on a white background. From the state of art in color research, we know that the immediate environment of colors influences their perception [41, 62, 141]. Different background choices have a particularly strong effect. Background color interference is therefore a further area of investigation.

10. Conclusion

We conducted two studies with similar results in which participants identified the color differences between two nodes in a node-link diagram. We varied link topologies, link hues, and node hues. Our results show that node colors are significantly easier to discriminate when a link separates the space between them, regardless of the link hue. Links that connect the two nodes decrease discriminability. Furthermore, we found that

complementary colored links improve the discriminability of the node colors, regardless of the topology. Links with a similar hue to the node colors hinder the discriminability. The color coding of nodes that we explored was particularly well suited for representing numerical data. Based on our results, we formulated guidelines for the choice of colors for nodes and links in node-link diagrams to improve visualization design. Specifically, based on our tested colors, we recommend using blue for nodes with the greatest brightness differences and gray links to maximize the distinction between node colors. Our study provides interesting and useful insights to the perception of color differences in node-link diagrams in comparison to scatterplots. These insights are useful for rapid initialization of node link diagrams and creating graphs for exploration across various application domains, such as medicine, biology, finance or social media.

11. Acknowledgments

We thank the study participants. We are grateful to the reviewers for their constructive comments. We would also like to thank D. Braun, M. Sondag, S. Bremm and T. Baumgartl for their support. This research was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [527250730] and Bundesministerium für Bildung und Forschung (BMBF) [01ZZ2323J].

12. Appendix

12.1 Further Analysis on Study 1

Axis	Link Topology/Color	Slope	R^2	ND(50%) ΔE	ND(50%) ΔE_{00}
L	Scatterplot	.087	.965	5.75	
L	None/None	.0902	.9024	5.5456	4.0755
L	Connecting/Gray	.0881	.9146	5.6750	4.1425
L	Separating/Gray	.0892	.8943	5.6033	4.1321
L	Connecting/Same	.0838	.9346	5.9642	4.3258
L	Separating/Same	.0883	.9111	5.6613	4.1402
L	Connecting/Complementary	.0864	.9235	5.7863	4.2141
L	Separating/Complementary	.0903	.9118	5.5353	4.0488
L	Crossing/Gray	.0859	.9148	5.8213	4.2515
L	Crossing/Same	.0860	.9210	5.8113	4.2370
L	Crossing/Complementary	.0853	.9236	5.8607	4.2613
a	Scatterplot	.064	.992	7.81	
a	None/None	.0654	.9587	7.6427	3.6992
a	Connecting/Gray	.0631	.9657	7.9238	3.8355
a	Separating/Gray	.0671	.9170	7.4557	3.6487
a	Connecting/Same	.0627	.9483	7.9708	3.8700
a	Separating/Same	.0649	.9573	7.6994	3.7511
a	Connecting/Complementary	.0648	.9375	7.7208	3.7683
a	Separating/Complementary	.0672	.9348	7.4382	3.6180
a	Crossing/Gray	.0659	.9549	7.5848	3.6854
a	Crossing/Same	.0578	.9707	8.6454	4.2166
a	Crossing/Complementary	.0638	.9496	7.8348	3.8074
b	Scatterplot	.05	.979	10.03	
b	None/None	.0535	.9634	9.3422	4.4738
b	Connecting/Gray	.0529	.9477	9.4564	4.5443
b	Separating/Gray	.0545	.9263	9.1806	4.4861
b	Connecting/Same	.0511	.9457	9.7846	4.7034
b	Separating/Same	.0556	.9676	8.9936	4.3163
b	Connecting/Complementary	.0515	.9643	9.7075	4.6698
b	Separating/Complementary	.0537	.9560	9.3106	4.4738
b	Crossing/Gray	.0541	.9636	9.2441	4.4194
b	Crossing/Same	.0478	.9507	10.4644	5.0187
b	Crossing/Complementary	.0524	.9522	9.5430	4.5825

TABLE 1: Mean ND(50%) in ΔE and ΔE_{00} for different link combinations added as stimuli per CIELab axis. This table presents a comparison of the values derived by Szafir [152].

12.2 Node Hues and Colors

	Blue L*,a*,b*	Green L*,a*,b*	Red L*,a*,b*	Yellow L*,a*,b*	Pink L*,a*,b*	Turquoise L*,a*,b*
Original	25,70,-90	75,-75,70	50,70,60	85,-20,90	50,95,-60	85,-45,-15
L-5	20,70,-90	70,-75,70	45,70,60	80,-20,90	45,95,-60	80,-45,-15
L-10	15,70,-90	65,-75,70	40,70,60	75,-20,90	40,95,-60	75,-45,-15
L-15	10,70,-90	60,-75,70	35,70,60	70,-20,90	35,95,-60	70,-45,-15
L+5	30,70,-90	80,-75,70	55,70,60	90,-20,90	55,95,-60	90,-45,-15
L+10	35,70,-90	85,-75,70	60,70,60	95,-20,90	60,95,-60	95,-45,-15
L+15	40,70,-90	90,-75,70	65,70,60	100,-20,90	65,95,-60	100,-45,-15
a-5	25,65,-90	75,-80,70	50,65,60	85,-25,90	50,90,-60	85,-50,-15
a-10	25,60,-90	75,-85,70	50,60,60	85,-30,90	50,85,-60	85,-55,-15
a-15	25,55,-90	75,-90,70	50,55,60	85,-35,90	50,80,-60	85,-60,-15
a+5	25,75,-90	75,-70,70	50,75,60	85,-15,90	50,100,-60	85,-40,-15
a+10	25,80,-90	75,-65,70	50,80,60	85,-10,90	50,105,-60	85,-35,-15
a+15	25,85,-90	75,-60,70	50,85,60	85,-5,90	50,110,-60	85,-30,-15
b-5	25,70,-95	75,-75,65	50,70,55	85,-20,85	50,95,-65	85,-45,-20
b-10	25,70,-100	75,-75,60	50,70,50	85,-20,80	50,95,-70	85,-45,-25
b-15	25,70,-105	75,-75,55	50,70,45	85,-20,75	50,95,-75	85,-45,-30
b+5	25,70,-85	75,-75,75	50,70,65	85,-20,95	50,95,-55	85,-45,-10
b+10	25,70,-80	75,-75,80	50,70,70	85,-20,100	50,95,-50	85,-45,-5
b+15	25,70,-75	75,-75,85	50,70,75	85,-20,105	50,95,-45	85,-45,0

TABLE 2: Shows the color values and variations used in our experiments

12.3 Node-Link Diagram Coordinates

Nodes	x-coordinate	y-coordinate	Links	source	target
0	603	38	0	2	3
1	1148	761	1	5	7
2	969	379	2	0	7
3	1165	229	3	11	18
4	302	725	4	13	16
5	414	482	5	5	4
6	971	45	6	17	12
7	345	153	7	18	7
8	134	609	8	18	17
9	115	712	9	6	11
10	602	359	10	20	2
11	253	330	11	9	8
12	34	506	12	6	14
13	822	775	13	14	0
14	798	351	14	6	21
15	624	597	15	6	3
16	943	702	16	0	6
17	29	362	17	4	9
18	85	224	18	4	13
19	802	482	19	5	11
20	1010	578	20	4	20
21	1162	625	21	8	12
			22	15	12
			23	12	11
			24	5	10
			25	19	2
			26	1	13
			27	20	3
			28	20	15
			29	21	16
			30	1	16
			31	14	10
			32	14	2
			33	20	21

TABLE 3: Shows the node coordinates as well as the source and target nodes of the node-link diagrams used in our experiments

2.2 Discriminability of Node Colors in Node-Link Diagrams: Node Outlines

2.2.1 Introduction

The following paper examines the effects of outline colors, widths, and directions on the discriminability of node colors. It presents a series of studies that systematically vary outline designs while displaying nodes with different color differences, which study participants must identify. Just noticeable differences are determined to illustrate the influence of the outlines on the discriminability of node colors.

2.2.2 Visual Highlighting

Visual highlighting is a visualization technique that directs the viewer's attention to selected elements of interest without high cognitive effort required for a visual search. Highlighting uses modifications of one or more visual variables, such as color, size, opacity, lightness, or line width, to increase visual contrast of the selected objects against their surroundings [168]. Effective highlighting uses preattentive visual processing, in order to enable users to quickly detect important elements without the need for further inspections [60].

In node-link diagrams, highlighting is usually used to emphasize nodes, paths or clusters. This makes tasks such as target detection, path tracing, and structural exploration easier. However, the effectiveness of highlighting depends on the discriminability of the visual encodings [108, 168].

2.2.3 Discriminability of Node Colors in Node-Link Diagrams and the Influence of Node Outlines

This paper is currently under review at *Information Visualization*.

I am the primary author of this publication. In this role, I was responsible for the design, implementation, data collection and analysis, as well as the writing and publication of the work. The individual contributions of myself and my co-authors to this publication are outlined below according to the Contributor Roles Taxonomy (CRediT):

Laura Pelchmann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft.

Julia Wieja: Julia conducted her master's thesis on the impact of node outlines on the discriminability of node colors, guided by me and supervised by Tatiana. Her research serves as the pilot study in this paper. Moreover, she proofread the final paper submission. Contributor Roles: Data curation, Formal analysis, Investigation, Software, Writing – review & editing.

Rita Borgo: Rita has made important contributions through her expertise in study design, implementation, and statistical analysis. Moreover, she was always available to provide feedback and answer questions throughout every aspect of the research process. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Margit Pohl: Margit has given important insights into psychological studies and their design, pertinent previous literature, and the appropriate variation of stimuli. Additionally, she consistently provided feedback and addressed questions about all aspects of the research process. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Tatiana von Landesberger: Tatiana has extensive expertise in designing graph comparison studies and has been instrumental in supporting me throughout the research project, offering crucial feedback and constantly aiding in decision-making processes. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

The Impact of Node Outlines on the Recognition of Color Differences in Node-Link Diagrams

LAURA PELCHMANN¹, JULIA WIEJA¹, RITA BORGO², MARGIT POHL³, TATIANA VON
LANDESBERGER¹

¹ University of Cologne

² King's College London

³ TU Wien

Abstract:

Node-link diagrams are widely used to visualize graph data, with node colors serving as an effective method for representing quantitative node attributes. Distinguishing those node colors can however be challenging, especially when the colors are similar. Previous studies have shown that link colors can influence color discrimination, but the impact of node outlines as a highlighting technique has received little systematic attention. In this paper, we present two online experiments that vary the direction, color and width of node outlines to investigate their influence on users' ability to distinguish node colors in quantitative data visualization. To obtain results that are independent of the data-dependent quantities of highlights, we vary the amount of highlighted nodes across different studies. Our findings reveal that node outlines can significantly affect color discriminability of node colors, and should therefore be selected with great care. Based on our results, we offer practical guidelines for improving the color discriminability of node colors in node-link diagrams: Highlight only the most relevant nodes with outlines. Use outward-directed outlines. Prefer black outlines or yellow/orange outlines for blue nodes while avoiding red node outlines. Keep the outline thickness minimal. By following these guidelines, users can create node-link diagrams where node colors are more discriminable and thus more accurately represent the values.

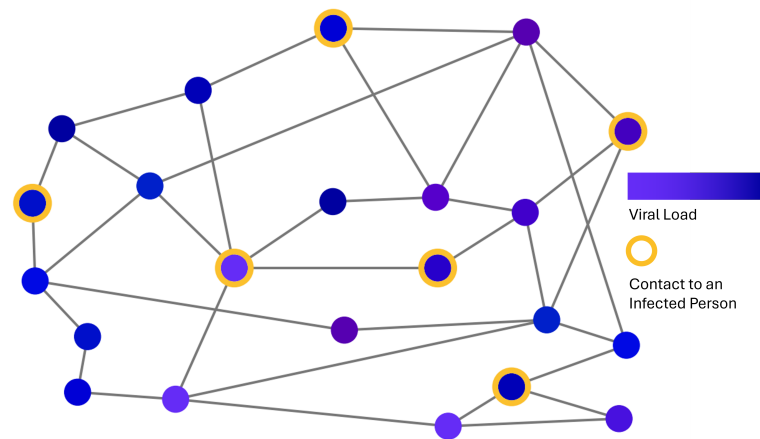


FIGURE 1: Shows an example of a node-link diagram with six nodes highlighted by node outlines. For these outlines, a width of 5 pixels was chosen in combination with a complementary color, and the outlines were directed outward.

1. Introduction

Color encoding of nodes is an effective method for visually representing various types of quantitative data in node-link diagrams, such as indicating infection risks in contact networks. Distinguishing those node colors is crucial for the accurate display of these data, especially when colors are similar or nodes are densely arranged. Highlighting specific nodes in addition to their color can improve the visibility of important data aspects (see Figure 1).

Colored outlines are an option for highlighting nodes without compromising the data faithfulness [88, 92, 124, 137, 138, 145, 164]. Since previous studies have shown that link colors can both interfere with or improve node color discrimination in node-link diagrams [114], understanding the role of additional visual attributes, such as node outlines, is important for accurate data readability, when using them for highlighting (see Figure 1).

To examine the impact of node outlines on the perception of node color differences [88, 92], we conducted two experiments comprising nine online studies. For these experiments, we built upon previous studies conducted by Pelchmann et al. [114], who investigated the effect of link colors on the color discriminability of node colors. Our new studies aimed to explore how different node outlines affect users' ability to distinguish node colors in quantitative color schemes.

For this purpose, we systematically varied the direction, color and width of node outlines within each study. Participants were to compare node colors with different outline designs to measure the effect of each factor on the color discrimination.

Between the studies we varied different amounts of highlighted nodes, representing potentially significant data aspects, in order to emphasize the generalizability of the experiments.

Our findings provide a basic understanding of how node outlines influence the perception of color differences in node-link diagrams, identifying how node outlines can support both perceptual accuracy and expressive visualization design.

As a result we provide specific design guidelines for our tested stimulus selections to support users' decisions on the use of outlines to highlight nodes without distorting the underlying quantitative color encoding, such as:

- Highlight only the most relevant nodes with outlines.
- Use outward-directed outlines.
- Prefer black outlines or yellow/orange outlines for blue nodes while avoiding red node outlines.
- Keep the outline thickness minimal.

With the help of these guidelines, both expert and novice users can utilize node outlines when creating node-link diagrams to effectively highlight certain data aspects of quantitative data structures that are represented by the node color. For example, to highlight infected individuals in medical networks (see [Figure 1](#)).

In the following, we first examine relevant works from previous research. Subsequently, we formulate the resulting research questions and corresponding hypotheses. After explaining the experimental procedure, we discuss the results, the resulting guidelines, and potential future work. Finally, we end with our conclusion.

2. Related Work

First, we discuss the importance of color as a visual encoding in node-link diagrams. Color is commonly used to represent both categorical and quantitative attributes in node-link diagrams, playing a central role in enabling users to differentiate, compare, and interpret the underlying data [44, 77, 114]. We review prior research on just noticeable differences (JNDs) in color perception. This research provides the perceptual foundation for understanding how small color variations are detected and how reliably quantitative values can be distinguished. We then investigate the role of color in node-link diagrams. Lastly, we introduce related work on highlighting in visualizations, specifically in node-link diagrams.

Color in Visualization Using color for visualizations is an important aspect of research in the visualization community [168]. The key research questions concentrate on how colors can effectively represent data and how to effectively assign colors to data [36, 91, 98]. Accurately mapping different numerical values to colors provides a specific problem, since commonly one hue is used, only varying the saturation of the used colors [26, 95, 127]. Hence, color schemes are needed that use discriminable colors, for varying data types and numbers of data values. Visualization researchers have developed multiple tools to find effective color schemes depending on the data being visualized [54, 57, 106, 154, 167]. But, these tools cannot guarantee an accurate representation of both values and attributes.

Studies on color discriminability often use predefined color sets but may not account for factors like background color, nearby areas [102, 141, 170, 172] or contrasting colors [33, 80, 129, 169]. We examine how different node outlines might affect discriminability of node colors in node-link diagrams. We then compare these results to previous studies on color discriminability in scatterplots. Comparing nodes in node-link diagrams to points in scatterplots, while adding edges, is inspired by [9]. A lot of previous research focuses on the improvement of categorical color schemes [158, 167], while our work aligns with Szafir's focus on quantitative visualization [152]. While most experiments are conducted under controlled conditions, which may limit real-world applicability (for example the staircase method [37]), our methodology is inspired by Szafir [152] and aims to address these limitations.

Node-Link Diagrams Studies examining the distinctions in color within graph drawing remain insufficiently explored. Specifically, the role of color in node-link diagrams is a topic that has received limited attention [55]. Burch et al. [27] identified the use of color in node-link diagrams as a notable challenge. The visual attributes of design choices in node-link diagrams significantly impact their interpretation [101] and therefore human comprehension of the data and its complexity [122, 149, 173]. Saket et al. assessed the effectiveness of visualizing relationships in node diagrams, node-link diagrams, and node-link-group diagrams [134]. Their findings indicate that incorporating links or both links and group representations does not diminish performance on node-based tasks. A lot of perceptual research on node link diagrams focuses on the effect of different link designs on data comprehension. For instance, the impact of rounded links [68] and the effect of link crossings [69] have been explored. Reimann et al. [128] investigate how colored links in clustered node-link diagrams can enhance cluster differentiation. Their work focuses on the effectiveness of colored links in distinguishing groups of individuals, but does not examine the color discriminability of the nodes. All of these studies suggest that colored node outlines also have a significant impact on the discrimination of node colors and therefore inspire our research.

Highlighting in Visualization Color is also a key visual channel for highlighting elements in visualizations, but in multi-class scatterplots, it often reduces color discriminability [96]. In this paper, Lu et al. propose a context-preserving highlighting method that maintains perceptual separability and consistent color mapping during interactive exploration, using two contrastive color schemes combined dynamically for effective highlighting. Waldner et al. [164] focus their investigation on multi-channel highlighting in scatterplots, using luminance and blur. They find that this combination leads to a good trade-off between highlight effectiveness and aesthetics.

The choice of color is not the only factor that affects the readability of a visualization. Cleveland and McGill [36] demonstrated that different color channels convey individual values less precisely than the visual variable size. To change the size of a point, various approaches can be used. The point can be enlarged by increasing the colored radius, or it can be highlighted using an outline. If the outline is outside the point, the point as a whole is enlarged.

If the outline is inside the point, the colored area shrinks. As the point size decreases, comparing colors becomes increasingly difficult, as Szafir [152] demonstrated in her study on scatterplots. This work examines whether this phenomenon also occurs in the context of node-link diagrams.

A study by Smart et al. [148] demonstrated that the perception of color differences in a scatterplot, where objects are colored in a quantitative color scale, is significantly influenced when the shape is changed. In their study, both filled shapes and unfilled symbols (i.e., shapes consisting only of the outer line) were examined. The participants had greater difficulty differentiating the colors of the unfilled shapes compared to their fully colored counterparts. Since the unfilled shape can be considered an outline, the question arises whether the perception of color differences is further hindered when this unfilled shape is placed around its filled counterpart.

3. Research Question and Hypotheses

Following these previous investigations, the key questions for our research are:

What directions, colors and widths should be used for node outlines when highlighting specific nodes in node-link diagrams?

Should these directions, colors and widths be chosen differently depending on the amount of highlighted nodes?

To answer these questions we developed the following, more specific research questions and corresponding hypotheses:

RQ1: What influence does the amount of highlighted nodes with outlines in node-link diagrams have on the perception of differences between node colors?

H1: The more highlighted nodes with colored outlines are in a node-link diagram, the more difficult it is to distinguish the node colors.

RQ2: What influence does the direction of node outlines in node-link diagrams have on the perception of color differences between nodes?

H2: Distinguishing node colors is more difficult with inward directed outlines than with outward directed outlines.

RQ3: What influence does the color of node outlines in node-link diagrams have on the perception of color differences between nodes?

H3: The use of red color for outlines most impairs the differentiation of node colors, while black outlines have the least negative impact on the color distinction of the node colors.

RQ4: What influence does the width of node outlines in node-link diagrams have on the perception of color differences between nodes?

H4: The thicker the node outline, the more difficult it is to distinguish the node colors.

4. Experiment Design

To answer these research questions, we are building on the study design used by Pelchmann et al. [114]. In those studies, the effect of edge colors and topologies was tested; in this new experiment, we are modifying the stimuli to investigate the effect of node outlines on the perception of differences in node colors.

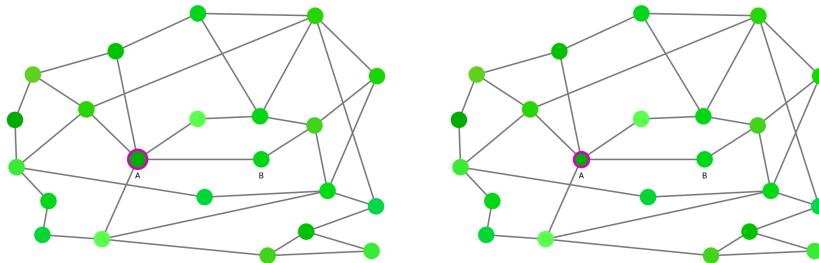
We first conducted a preliminary experiment in order to test the settings with one highlighted node before conducting the main experiment.

4.1 Study Variables

The detection of color differences serves as the dependent variable in both studies. In order to test the differentiability, we manipulate the *Amount of Outlined Nodes*, *Outline Direction*, and *Outline Color*, as well as the *Outline Width*, which serve as our independent variables. To avoid overwhelming the participants with too many trials and to ensure a valid data basis, the studies were designed as between-subject studies, with half of the participants comparing nodes with outlines directed outward and the other half comparing nodes with outlines directed inward. The variables *Outline Color* and *Outline Width* were examined within the studies.

Preliminary Experiment For the preliminary experiment we investigated node-link diagrams with green nodes, varying between pink, gray and black outlines in combination with thin, middle and thick outline widths. The two studies, which were conducted in parallel, differed in terms of the outline direction:

- One highlighted node with outward-oriented outline (left), one with inward-oriented outline (right):

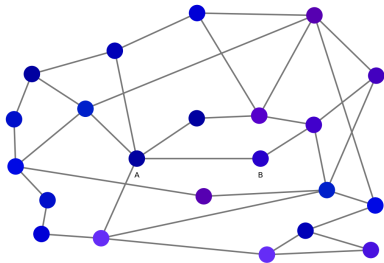


Main Experiment Based on the findings of the preliminary experiment, we developed a final setup adding a variation for the *Amount of Outlined Nodes* additionally to the *Outline Direction* as a between subject study. To ensure comparability with previous experiments conducted without outlines [114], a study without highlights was also performed. This study maintained the same setup and number of trials as the others.

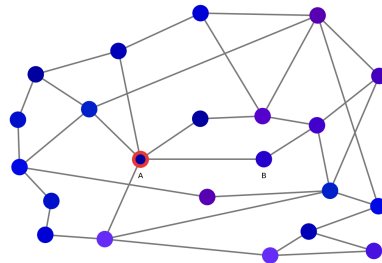
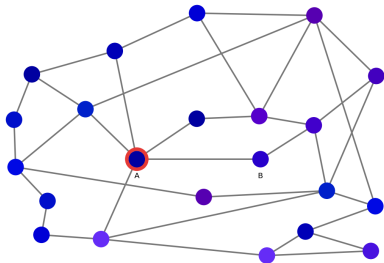
Since the results from Pelchmann et al. [114] suggest that using blue nodes for color discrimination yields better results, we changed the color of the nodes to blue for the main study. Furthermore, since the results showed that gray outlines had little highlighting effect, we adjusted the main experiment to test black, yellow/orange, and red outlines in combination with thin, middle and thick outline widths.

This resulted in the following seven studies, displaying node-link diagrams with blue nodes, which were conducted in parallel, differing in terms of the outline direction and amount of highlighted nodes:

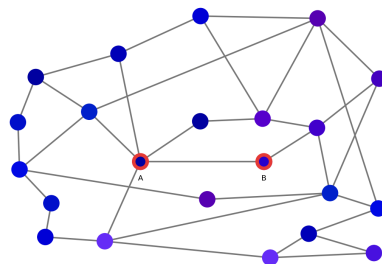
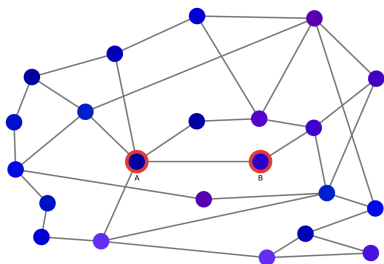
- No highlighted nodes:



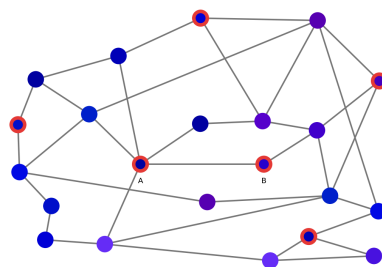
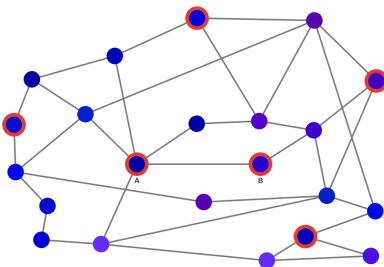
- One highlighted node with an outward-oriented outline (left) and an inward-oriented outline (right):



- Two highlighted nodes with an outward-oriented outline (left) and an inward-oriented outline (right):



- Six highlighted nodes with an outward-oriented outline (left), and an inward-oriented outline (right):



This multi-experiment design allowed us to systematically examine how outline direction, outline color, and outline width influence color discrimination in node-link diagrams with various numbers of highlighted nodes.

4.2 Task

In both studies, participants were shown images of the same node-link diagram for which they had to decide whether two specific nodes A and B in the center of the node-link diagram had the same color or not (see the screenshot in [Figure 2](#)).

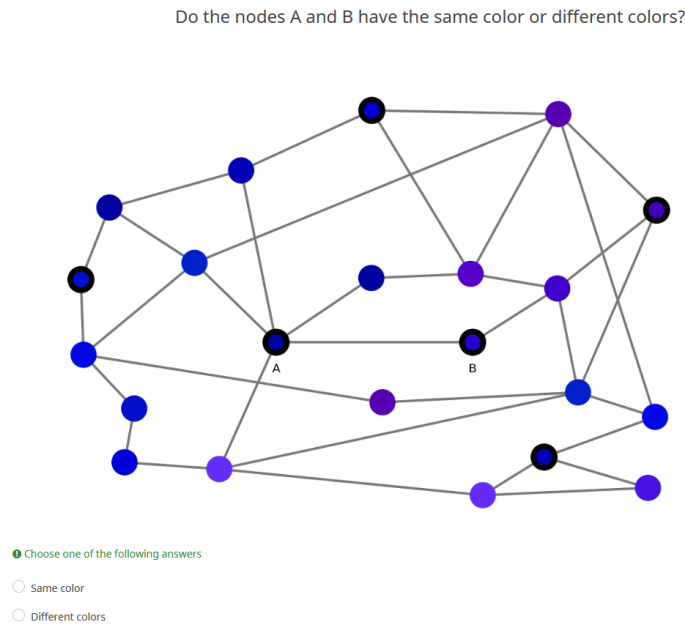


FIGURE 2: Shows an exemplary screenshot from the study with six highlighted nodes and black, five pixel outward oriented outlines.

4.3 Stimuli

4.3.1 Node-Link Diagram

Following Pelchmann et al. [114], we designed a node-link diagram that enables the comparison of results with Szafir's scatterplot experiments [152]. The diagrams were rendered at a size of 1205×805 pixels, with the nodes drawn within an area of 1136×787 pixels. The background was left in white. The edges connecting the nodes were plotted with a width of three pixels. To avoid potential conflicts with the node colors, they were colored in a medium gray tone (in CIE Lab: $L^* = 50$, $a^* = b^* = 0$, HEX: #777777) [93, 152]. Each diagram included 22 nodes and 35 edges. The graph represented by this node-link diagram was undirected. The topology was defined in advance and kept constant throughout the entire study period. This approach served to avoid the confounding effect [148], which refers to the impact that confounding variables can have on dependent and independent variables.

The size of the nodes was also kept constant. All nodes had a size of 1000 pixels. Neither the nodes nor the edges overlapped other nodes. The two nodes to be compared were positioned horizontally centered in the diagram, each offset by 10% from the horizontal center of the graphic.

4.3.2 Node Colors

In both experiments, the colors of the two central nodes, A and B, should be compared. Node B remained in a constant color, while Node A varied in the displayed color. For the color variation, changes were made along each axis of the CIE Lab color space (ΔL^* , Δa^* , and Δb^*), with changes of ± 7 and ± 15 . This resulted in 13 different color values that Node A could be displayed in. The remaining nodes in the node-link diagram were randomly colored in one of these 13 color values, which then remained constant and did not change further.

Preliminary Experiment The color chosen for node B in the preliminary experiment was a light green, this color corresponds to the string #00d619. The variations of this color were within a range of $L^* = [60, 90]$, $a^* = [-90, -60]$ and $b^* = [55, 85]$, comparable to the scatterplots used in [152].

Main Experiment The color chosen for node B in the main experiment was a medium blue, this color corresponds to the string #2700CB. The variations of this color were within a range of $L^* = [10, 40]$, $a^* = [55, 85]$ and $b^* = [-105, -75]$, comparable to the node-link diagrams used in [114].

4.3.3 Outline Directions

For the positioning of the outline, there were two different options. The outline could either lie inside the nodes, ensuring that the overall size of these nodes remained consistently at 1000 pixels, regardless of the outline's width. In this case, the colored area inside of the outlined nodes was reduced according to the outline's width. The other option was that the outline lay outside the node. Hence, the colored area of these nodes remained at 1000 pixels, and the outline increased the overall size according to its width.

4.3.4 Outline Colors

Preliminary Experiment Three different outline colors were chosen for the preliminary experiment, all three chosen to convey specific information: black (#000000) was used to subtly indicate nodes of particular interest without drawing excessive attention to them, while a medium gray (#777777) matched the edge color, maintaining visual consistency. Alternatively, a complementary color to node B was chosen which corresponds to the HEX code #D600BE. Previous studies have shown that using complementary colors in the surroundings has a less detrimental effect on the discriminability of node colors than the node colors themselves [114]. Since the gray outlines in the preliminary experiment had the least effect on the perception of color differences while the highlighting effect was low, we decided to replace them with a color with a greater highlighting potential for the main experiment.

Main Experiment For the main experiment, the colors of the outlines were adjusted based on the color changes of the nodes. We tested black outlines (#000000) again to subtly indicate nodes of particular interest without drawing excessive attention to them, as well as complementary colored outlines; for the blue nodes in these studies we used yellow/orange outlines (#FBC02D) accordingly. The yellow/orange was chosen to be slightly darker than the yellow edges in the experiments conducted by Pelchmann et al. [114], as the outline should provide sufficient contrast to the background. To include a color with a high signaling effect for the highlighted node outlines, we chose red (#E53935) as the third tested outline color [164].

4.3.5 Outline Widths

The outline widths were consistent through both experiments, chosen in increments of two pixels: 1px, 3px or 5px wide. A thicker outline would exceed the size of the nodes in comparison, which is why testing beyond 5 pixels was not conducted.

4.4 Procedure

The two experiments had an identical structure. Participants were shown a node-link diagram and asked to indicate whether the colors of nodes A and B were the "same color" or "different colors."

After a brief introduction to the survey, participants were asked to complete an Ishihara test to exclude any color blindness. Following this test, participants were presented with a training session, where they viewed a node-link diagram with the same topology as the survey. The node colors came from a categorical color scale to illustrate the discriminability of the two nodes. In the training, the same question had to be answered as in the main part of the survey, with node A colored green and node B colored red.

After the training, the main survey began, participants conducted a total of 117 comparisons (13 different colors of node A $\times$ 3 different outline thicknesses $\times$ 3 different outline colors). After every question a white screen was shown to minimize visual interferences between the shown colors. After 56 comparisons, participants had the option to take a short break. The response options were always predefined and could be either "same color" or "different colors." There was no option to skip questions or refuse to answer. Responses were selected by clicking on the respective option. Participants could change their minds as often as they wanted. However, once they had made a decision and clicked "Next" to load the next visualization, they could not undo their last decision. Each of the 117 diagrams was displayed randomly, making it impossible to infer any sorting. Randomization resulted in each study participant having a different order.

After completing the comparisons, participants were asked to voluntarily provide the following personal information: age, gender, color blindness, prior knowledge about color differences, prior knowledge about graphs, and native language.

4.5 Participants

Preliminary Experiment In the preliminary experiment, responses from 106 participants from the university environment were considered. Responses from participants who failed the Ishihara test were excluded.

Main Experiment For the main experiment, our power analysis ($\alpha = 0.05$, power: $1 - \beta = 0.80$, effect: $f = 0.25$) indicated that we needed 64 participants per amount of highlighted nodes and per outline direction. These participants were acquired through Prolific. Participants had to speak English and could not have any color vision impairment (the Ishihara test was still conducted). Additionally, participants could not take part in multiple studies, and the group of previous participants was excluded. Thus, the responses of 64 to 66 participants per study were included in the evaluation after they passed the Ishihara test, resulting in a total amount of 454 participants for the main experiment.

5. Methodology

For our analysis we followed the calculations of Szafir [152] and Pelchmann et al. [114]:

1. **Normal distribution:** For each of our studies, the Kolmogorov-Smirnov test (sign. < .001) as well as the Shapiro-Wilk test (sign. < .001) rejected the null hypothesis that the data were normally distributed.
2. **Independence:** For every research question, the Kruskal-Wallis test was conducted to test the independence of the variables that were under investigation.

3. **Discriminability rate:** First, the discriminability rate was calculated for each of the 117 trials, by the following formula:

$$DRate = \frac{\text{amount of detected differences}(\Delta E)}{\text{amount of total trials}(\Delta E)}.$$

4. **ND(50%):** With the resulting discriminability rate per color difference, a linear regression was calculated. Following the previous works [114, 152], this regression line was forced through the origin, because the amount of people detecting differences between two identical colors should be zero. For this we used the following formula: $m_{axis} = \frac{\Delta E * DRate(\Delta E)}{DRate^2(\Delta E)}$.

We then used the slope of the linear regression line to determine the estimator for noticeable differences per axis: $ND_{axis}(p) = \frac{p}{m_{axis}}$.

For this p is the desired factor of the discriminability rate, i.e., the proportion of people spotted the difference between two colors displayed. Usually $p = 0.5$ is used for the investigation of just noticeable differences [114, 152]. Hence, the displayed ND(50%) values show the calculated results.

In [114], the mathematically improved distance of ΔE_{00} was additionally calculated and displayed. Although these results slightly deviated from the calculated values of ΔE , they did not affect the calculated order of the complicating factors in any case of the color discriminability. Therefore, we decided to use only the ΔE values for this calculation in the interest of simplified presentation.

6. Results

6.1 Preliminary Experiment

The preliminary experiment provided the following insights, based on which we developed the subsequent experiments:

- Directing node outlines outwards improves the perception of color differences of node colors compared to directing them inwards.
- Thicker outlines reduce the perception of color differences in node colors compared to thinner outlines.
- Gray outlines have the least effect on color difference perception of node colors in comparison to black and complementary colored outlines. Black outlines improve the perception of color differences of node colors in comparison to complementary colored outlines.

6.2 Main Experiment

For the evaluation of the results, we focus on the four hypotheses we have established, which examine each of our stimuli.

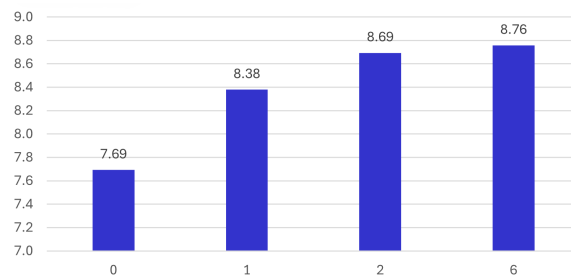


FIGURE 3: Shows the average of ND(50%) in ΔE for the tested amount of highlighted nodes in the main experiment.

Number of Outlined Nodes The results confirm our hypothesis **H1**: The more highlighted nodes with colored outlines are in a node-link diagram, the more difficult it is to distinguish the node colors (see Figure 3).

This result highlights the importance of our investigation, as the required color difference of the nodes in node-link diagrams is significantly lower in the experiment without highlights compared to the experiments with highlighted nodes. The required color difference increases with a higher number of color-outlined nodes. The number of nodes to be highlighted is naturally determined by the data, which is why this result particularly motivates further aspects of the investigated influencing factors.

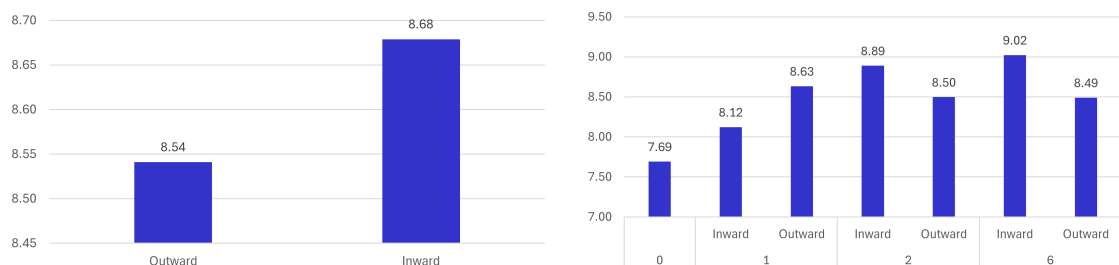
The test for independence between the tested amount of outlined nodes showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = < .001, rejecting the null hypothesis of dependent categories.

Outline Directions The results confirm our hypothesis **H2**: Distinguishing node colors is more difficult with inward directed outlines than with outward directed outlines.

This effect is evident for the results over all studies (see Figure 4a) and in those with two six highlighted nodes. Only with one node highlighted by a colored outline do the results show easier discriminability of node colors when the outline is directed inward (see Figure 4b).

With only one highlighted node, the size change in just one node, due to the outward-directed outline, might have a greater negative impact on discriminability. However, this phenomenon remains to be investigated in further studies.

The test for independence between the tested outline directions showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = < .001, rejecting the null hypothesis of dependent categories.



(A) Shows the average of ND(50%) in ΔE for the tested outline directions in the main experiment.

(B) Shows the average of ND(50%) in ΔE for the tested outline directions depending on the amount of highlighted nodes in the main experiment.

FIGURE 4

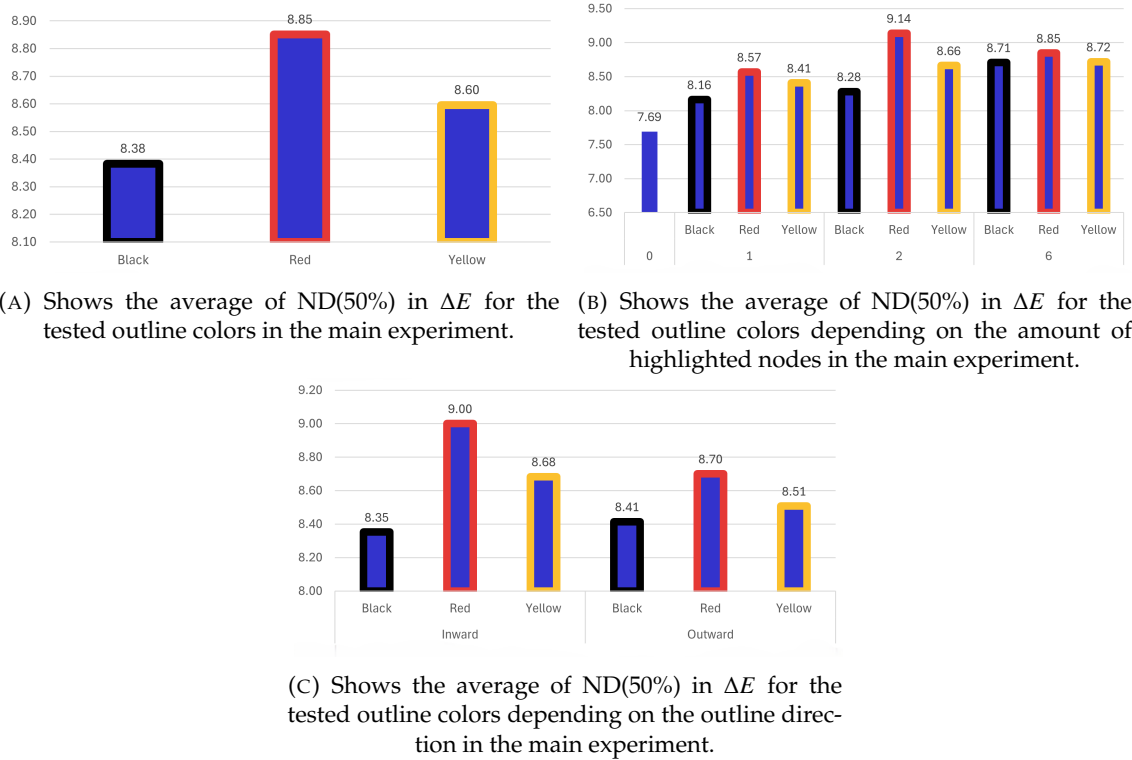


FIGURE 5

Outline Colors The results confirm our hypothesis **H3**: The use of red color for outlines most impairs the differentiation of node colors, while black outlines have the least negative impact on the color distinction of the node colors.

This is evident over all studies (see Figure 5a) and in dependence on the amount of highlighted nodes (see Figure 5b). However, it shows that the difference in the negative effect is significantly lower with six highlighted nodes than with one or two highlighted nodes. Regarding the direction of the outlines (see Figure 5c), the results also support our hypothesis **H3**.

These results support findings from previous studies on the color influence of surrounding areas, such as simultaneous contrast [41]. Due to the darkness the black outline is most similar to the blue color of the nodes and thus has the least impact on the color discriminability of the node colors, while the yellow/orange outline, due to its lighter color, forms a lower contrast with the white background and thus does not negatively affect the color discriminability of the node color as strongly as the red outlines, which have a high highlighting effect.

The test for independence between the tested outline colors showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = < .001, rejecting the null hypothesis of dependent categories.

Outline Widths The results confirm our hypothesis **H4**: The thicker the node outline, the more difficult it is to distinguish the node colors (see Figure 6a).

This result is evident in all three, the dependence on the amount of highlighted nodes (see Figure 6b), across the outline directions (see Figure 6c) and various outline colors (see Figure 6d). This clearly shows that thick outlines have a strongly negative impact

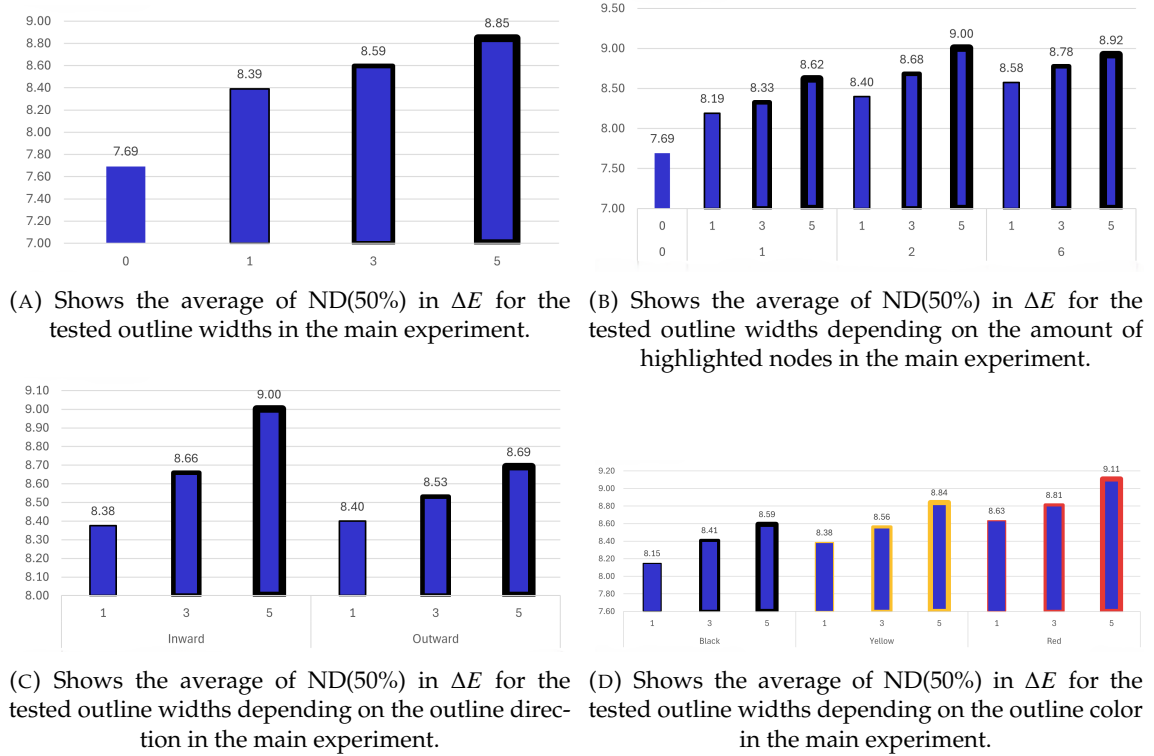


FIGURE 6

on the color distinction of node colors. Therefore, for highlighting with colored outlines, line widths of 1 to a maximum of 3 pixels should be used, depending on how strong the desired highlighting effect needs to be.

The test for independence between the tested outline widths showed a significant effect. Kruskal-Wallis, Asymptotic Significance (2- sided test) = < .001, rejecting the null hypothesis of dependent categories.

7. Guidelines

Based on the results for our tested stimuli, the following user guidelines can be formulated for the use of node outlines for highlighting purposes:

- Reduce the amount of highlighted nodes with outlines in a node-link diagram as far as possible to maintain clarity in node color discriminability.
- Use outward directed outlines instead of inward directed outlines for small to medium sized nodes to make it easier for users to distinguish node colors.
- Black outlines have the least negative impact on color discriminability of blue nodes and are therefore a high quality option when outlining nodes.
- If black outlines do not sufficiently meet the highlighting requirements, yellow/orange node outlines do not impair the differentiation of node colors for blue nodes as much as red outlines, while still maintaining strong highlighting purposes.
- Use thinner outlines to make it easier to distinguish node colors. Avoid thick outlines as they can impede the node color discriminability.

8. Limitations and Future Work

Our research focuses on first insights into the factors of node outlines that influence the perception of color differences of nodes. The node colors represent quantitative node attributes, the same colors were previously used in [114]. Color variations, for example, in the edges, are disregarded for now in order to identify initial implications that can serve as a foundation for future work.

The layout we chose for the node-link diagram shown here is based on a force-directed algorithm. This is a widely used layout algorithm that has yielded good results in previous studies regarding human performance with the data [121]. Other representations of the node-link diagrams could consequently lead to different findings and need to be investigated further.

As shown by Pelchmann et al. [114], different edge colors influence the distinction between node colors, the use of differently colored edges in node-link diagrams to be colored could result in interference.

The results for outline direction suggest that outward directed outlines make it easier to distinguish colors than inward directed ones. In our study, we used nodes of a smaller to medium size, for larger nodes, this result may differ. The outline widths examined can also be classified only in terms of the investigated node size.

The guidelines for outline colors are based on the node color; for more comprehensive guidelines, different node colors, as well as additional outline colors and their various combinations, must be tested.

Since surrounding colors have a direct influence on the perception of colors [41, 62, 141] our results and guidelines could differ for other background colors. The node-link diagrams used in our studies were shown on white background, highlighting nodes on a darker background, for example in a dark-mode, could require different outline colors and would therefore be an interesting subject for future studies.

Future work could also focus on examining the effects of different node sizes: analyzing these node sizes in relation to both edge width and outline width could provide interesting insights for the future creation of node-link diagrams.

When visualizing categorical datasets using node colors, a particular challenge lies in finding suitable outline colors, since the nodes already have different hues. Research into effective methods for highlighting nodes in node-link diagrams that represent categorical datasets and how they differ from scatterplots, is therefore a particularly desirable focus for future research.

9. Conclusion

We conducted two experiments including nine studies with consistent results where participants had to identify color differences between two nodes in node-link diagrams, building up on the work of Pelchmann et al. [114]. Instead of investigating colored edges, we used colored outlines for nodes in node-link diagrams to highlight specific nodes, varying outline direction, outline color as well as outline width and examined their effect on the node color discriminability. Our results indicate that for small to medium sized nodes the node colors are significantly more discriminable when the node outlines are directed outward. Black outlines have the least negative impact on the discriminability, while in cases of the need for a better highlighting effect, we advise to use yellow /orange outlines over red outlines for blue nodes. Furthermore, we found that thinner outlines improve the discriminability of the node colors, compared to thicker outlines, regardless of their color and direction. Based on these results, we formulated guidelines for selecting outline colors, widths, and directions for nodes in node-link diagrams to improve

visualization design. Our study provides valuable insights into the perception of differences of node colors in node-link diagrams compared to scatterplots. These insights are beneficial for the initialization of node-link diagrams for data exploration across various application domains, such as medicine, biology, finance, or social media.

10. Acknowledgements

We express our gratitude to all study participants. We appreciate the constructive feedback provided by the reviewers. This research was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [527250730] and the Bundesministerium für Bildung und Forschung (BMBF) [01ZZ2323].

2.3 Discriminability of Node Colors: Large Value Ranges

2.3.1 Introduction

The following paper examines the performance for various adapted versions of the OMC color scheme for large value ranges in multiple types of visualizations. In doing so, this work presents two improved versions of the original OMC color scheme, that provides better performance for various user tasks.

2.3.2 Large Value Ranges

Large value ranges (LVR) data sets are characterized by values spanning several exponents (i.e., orders of magnitude) as used for scientific notations [16, 70]. The concept of this notation, is that each value v is divided into the mantissa $m \in \mathbb{R} \setminus \{0\}$ and its exponent $e \in \mathbb{Z}$:

$$v = m \cdot 10^e.$$

For this work, the most common base 10 is used while employing the scientific normalized notation [6], in which e is chosen so that $|m| \in [1, 10)$. Large value ranges are a significant factor for many scientific fields (e.g., meteorology [139] or physics [175]). Therefore, an effective visualization design for large value ranges is an important part of visualization research.

2.3.3 Color Vision Deficiencies

Color vision deficiencies (CVDs) are conditions that reduce a person's ability to distinguish between certain colors. These deficiencies can be inherited or acquired, with red-green deficiencies being the most common form of CVD across the world. CVDs are caused by changes in the photopigments of the eye's long-wavelength or medium-wavelength cones. Approximately 8% of males and 0.4% of females of European Caucasians descent are affected by congenital red-green color vision deficiencies, whereas blue-yellow color vision deficiencies and complete color blindness are significantly less common [11]. These numbers make color accessibility an important aspect for the design of visualizations [11, 71]. Since viewers with CVDs experience a lower discriminability between certain colors, visualizations that rely solely on color to encode data attribute values may become difficult to read for some viewers, if not carefully considered. Therefore, visualization designers must use color palettes with high perceptual discriminability [71, 168].

2.3.4 Effective Color Schemes for Large Value Ranges Optimizing Color-Coding Criteria

This paper is currently in preparation for resubmission to *PacificVIS*.

I am the primary author of this publication. In this role, I was responsible for the design, data analysis, as well as the writing and publication of the work. The individual contributions of myself and my co-authors to this publication are outlined below according to the Contributor Roles Taxonomy ([CRedit](#)):

Laura Pelchmann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft.

Malin Böhme: Malin conducted her master’s thesis on effective color schemes for large value ranges, guided by Daniel and me and supervised by Tatiana. In this regard, she has notably contributed to the formulation and technical implementation of the guidelines derived from the literature, as well as the conduct and analysis of the evaluation. Moreover, she proofread the final paper submission. Contributor Roles: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing.

Daniel Braun: Daniel contributed important expertise in the visualization of large value ranges and the design and evaluation of OMC color schemes. He participated in the entire research process and wrote the initial draft of the evaluation section. Contributor Roles: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing.

Tatiana von Landesberger: Tatiana has extensive expertise in designing comparison studies and the visualization of large value ranges and has been instrumental in supporting throughout the research project, offering crucial feedback and constantly aiding in decision-making processes. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Effective Color Schemes for Large Value Ranges Optimizing Color-Coding Criteria

LAURA PELCHMANN¹, MALIN BÖHME¹, DANIEL BRAUN¹, TATIANA VON
LANDESBERGER¹

¹ *University of Cologne*

Abstract:

We present an optimization and automation process for generating color schemes to visualize data with Large Value Ranges (LVRs). Our novel color schemes apply common color quality criteria by improving the state-of-the-art Order of Magnitude Colors (OMC). We address OMC limitations in terms of hue selection, luminance adaptation, color-deficiency robustness, and adaptability to various number of exponents. We evaluate nine created color schemes both quantitatively via color metrics and through a human-subject study.

Evaluation provides evidence that two color schemes (optimized versions of the well-known glasgow color scheme) effectively improve the original OMC color scheme. Thereby, we provide designers with two effective color schemes for visualizing LVRs in various application domains, as demonstrated across diverse datasets and visualization types.

In preparation for resubmission at PacificVIS 2027

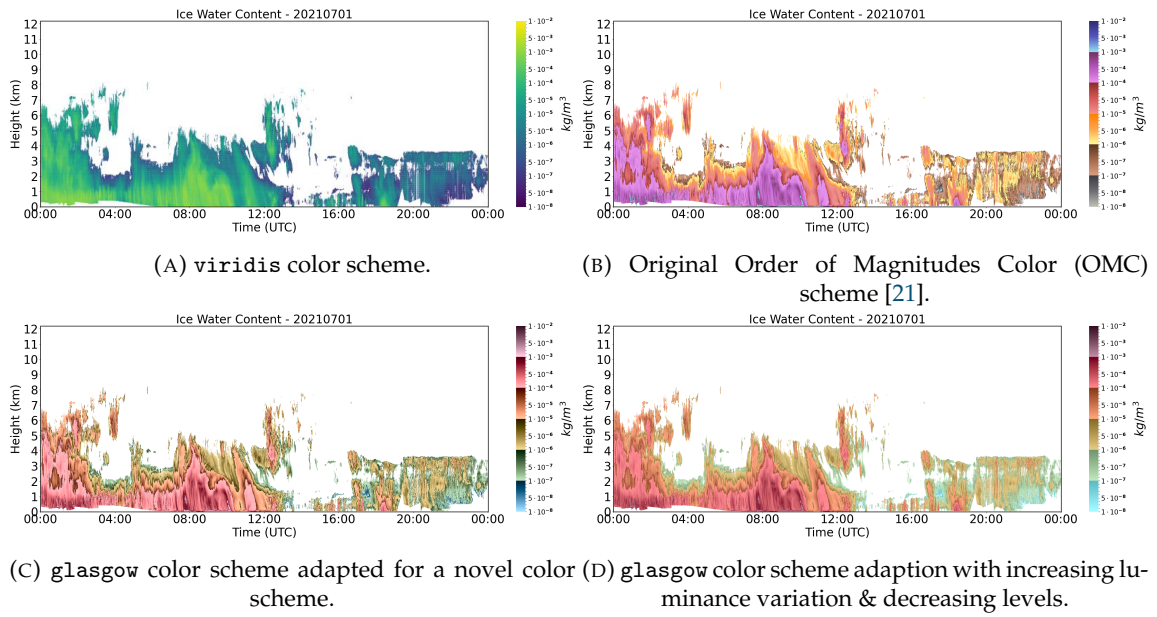


FIGURE 1: Color schemes applied to meteorological data displaying ice water content in a scatterplot using the original OMC scheme, and adapted versions based on *viridis* and *glasgow* and a second version of *glasgow* with increasing luminance variation.

1. Introduction

Color plays a central role in visualization, especially when the spatial position of data points is determined by the data itself, e.g., in scatterplots or node-link diagrams (see Figure 1). In these situations, color becomes a primary visual channel for distinguishing values and therefore supporting accurate data interpretation [109, 156]. But the effectiveness of color depends on the choice of a suitable color scheme. Only effective color schemes ensure accurate readability of values [21, 127]. The choice of color schemes is particularly difficult when encoding data with large value ranges (LVR) [6, 16, 21]. Data with large value ranges are datasets where the values span several different exponents when expressed in scientific notation ($v = m \cdot 10^e$) [21]. An example is a meteorological data set that displays measured particles of ice water content spanning six exponents (i.e., orders of magnitude). In case of LVR data, a large range of distinct values must be assigned to a limited number of colors, increasing the risk of misinterpretation [16, 132] (see Figure 1).

The Order of Magnitude Colors (OMC) scheme was introduced to represent data with LVR [21] (see Figure 1b). OMC separately encodes the exponent and the mantissa using hue variations and luminance gradients, respectively. This approach improves the value estimation and enables effective magnitude classification compared to commonly used sequential color schemes such as *viridis* [111]. Despite OMC's advantages over standard color schemes, the original OMC color scheme has considerable limitations. Specifically, hue selection has not been systematically adapted for maximal discriminability [109, 157], luminance gradients should be better adapted to varying value ranges per order of magnitude to improve perceptual order [2, 53, 89, 174], and accessibility for users with color vision deficiencies (CVD) has not been tested [103, 143, 156]. Furthermore, the original OMC color scheme only supports six orders of magnitude, while the input data might require a different number of exponents.

This work focuses on effectively improving and automating the creation of novel schemes, i.e., color schemes combining hue and luminance for color-coding data with LVR. Based on established criteria for effective color-coding, new color schemes are developed and enable an automatic adaptation to varying value ranges.

Our **contributions** are:

- We summarize existing color-coding criteria and adapt them to the special characteristics of large value ranges.
- We develop an algorithmic adaption process to optimize color-coding criteria to effectively improve existing color schemes.
- We evaluate the novel color schemes both quantitatively and empirically in a human-subject study.
- We exemplarily apply the novel color schemes to varying data and visualization types.

2. Related Work

Visualization of data with large value ranges presents unique challenges that require specialized techniques to ensure clarity and accuracy. A prevalent strategy [6,7,16,19,21,66] for visualizing large value ranges involves the distinct representation of the exponent e and the mantissa m of a value $v = m \cdot 10^e$. Nevertheless, the optimal mapping of these components to visual variables remains an open question. Our goal is to develop novel color schemes for these visualizations.

2.1 Visualization of Data with LVR

Hlawatsch et al. [65] pioneered the use of visualizations specifically tailored for large value ranges, which have been shown to surpass conventional methods in both accuracy and response time. Their adaption of the classic bar chart was further developed in recent works [16,66]. Other papers extended the approach of mantissa and exponent separation to time-dependent data [19,20] and general design considerations for LVR using split of mantissa and exponent [6,7,19]. They focus on visualization techniques, not on optimal color coding.

Braun et al. [21] investigated color-coding of large value ranges. They introduced the Order of Magnitude Colors (OMC) color scheme, where mantissa and exponent are encoded separately (see Figure 1b). They map exponent to hue/chroma and mantissa to luminance, ensuring non-overlapping perceptual channels following [155,157]. High spatial frequency is addressed through luminance variation [132,146]. However, OMC's hue as well as gradient selection does not sufficiently address common color-coding criteria.

2.2 Color-Coding Criteria

We review existing color-coding criteria and apply them to LVRs by adapting and improving existing color schemes. We categorize these criteria as general, specialized, and environment-related in order to easily extract the most relevant criteria for our use case. These criteria then form the basis for design decisions and define measurable targets for subsequent improvements.

General Criteria describe universally applicable principles for effective color-coding.

C1: Legend-based order ensures that colors are distinct, by selecting perceptually distant hues and by preventing duplicate colors across the luminance gradient variation [26, 89, 109, 157].

C2: Intuitive order requires that intermediate values can be perceptually inferred without consulting the legend, though this property may be task-dependent [132]. It is supported through monotonically increasing luminance (light-to-dark), presorted base hues, and sequential gradients [2, 57, 174], reinforced semantically via the dark-is-more bias [2, 53, 141, 174].

C3: Discriminative power is maximized by selecting distant hue selections and maximizing luminance variation within magnitudes while constraining it globally [15, 26, 89, 109, 118, 157].

C4: Uniformity demands proportionality between perceived color distance and underlying data distance [26, 89, 109]. Increasing uniformity intentionally raises **C3: discriminative power** across higher magnitudes via expanded luminance variation.

The widely used rainbow color scheme violates several of these principles, including **C2: intuitive order** and **C4: uniformity** and is therefore discouraged [2, 17, 26, 147].

C5: Color vision deficiency (CVD) robustness, requires distinguishability under color vision deficiencies to make visualizations accessible to a broad group of users [26, 103, 111, 143, 147, 156, 174]. **C6: Color name variation** quantifies the number of distinguishable and nameable colors and has been shown to support structure recognition and model comparison tasks [59, 127].

Specialized Criteria account for user, task, and data characteristics, such as *cultural background*, which influences semantic interpretation [143, 147, 156], and *aesthetics*, affecting user acceptance and interpretation [2, 26]. Aesthetics is often associated to higher chroma values [28] and smoothness, meaning avoiding artificial perceptual boundaries unless present in the data [89, 109]. Task-related criteria comprise *task type*, which determines appropriate color-codings [132, 146, 147, 156], and *highlighting*, where perceptual boundaries may be intentionally introduced [132, 146, 156]. Task-related criteria also favor segmented scales per magnitude, hue variation for value reading, sequential gradients for shape perception [53], and dataset-specific scale construction following the task-color-cube [156], while avoiding highlighting to preserve smoothness.

Data-related criteria include *semantics*, ensuring meaningful color associations [2, 174], and *emotions*, as color has been shown to dominate emotional response over data trend or visualization type [12]. *Data type* differentiates between ordered and unordered values [146, 147], while *unipolarity and bipolarity* motivate sequential versus diverging color schemes [57, 147]. Further considerations include *data properties* such as distribution and outliers [143, 156], *spatial frequency* describing local variation rates [132, 146, 147], and the *number of variables*, requiring distinct encoding channels in multivariate settings [155, 157].

Environment-Related Criteria are influenced by visualization type and environment. Visualization-related aspects include *spatial effects* in 3D rendering [146] and *color-size effects*, where color alters perceived size and vice versa [35, 152]. Environment-related criteria encompass *device gamut limitations* [143, 147, 156], *background effects* such as simultaneous contrast [141, 174], and *energy consumption*, which favors darker color schemes for sustainability and mobile usage [147, 174].

In sum, these criteria form a structured framework for evaluating and improving color-coding. In order to create color schemes with wide applicability, we focus on the General Criteria.

3. Color Scheme Creation

We use a bivariate color-coding for data with large value ranges $v = m \cdot 10^e$, inspired by Braun et. [21]: hue encoding the exponent and a luminance gradient encoding mantissa values. Our approach improves the selection of hues and the creation of suitable luminance gradient, both following the collected color-coding criteria (see previous section). We propose a multi-step approach:

1. *Hue selection*: We refine the selection of hues. We make them as distinct from each other as possible, while meeting all specified criteria. From the set of all possible improvements, the most promising options are selected by defining elimination conditions (i.e, filters and thresholds).
2. *Luminance gradient optimization*: We improve luminance gradients according to color-coding criteria.
3. *Combination and adaptation*: After combining hues and luminance gradients into color schemes, we present an algorithm for adapting color schemes to various number of exponents.

3.1 Hue Scheme Selection and Optimization

We present our approach to improve hue selections with respect to criteria relevant for hue selection: **C2**: *intuitive order*, **C3**: *discriminative power* [118], **C5**: *CVD robustness* and **C6**: *color name variation* [127]. This is done in several steps.

Initialization The initial hue selection is done using the state-of-the-art color selection tool `cols4all` [154]. We initialized the selection by choosing colorblind-friendly, categorical & sequential color schemes with 7 & 14 (only for sequential color schemes) colors. Seven and fourteen hues were taken, as the maximum number of distinctly namable colors is seven [59] – ensuring nameability. As they obey several criteria, both categorical and sequential color schemes were selected from the tool according to the following conditions.

- *Categorical* color-blind safe hue selection ensured **C5**: *CVD robustness*, **C3**: *discriminative power* and **C6**: *color name variation* criterion – thanks to the large perceptual distances between their colors [57]. Only categorical color schemes containing seven distinct colors were considered – for nameability.
- *Sequential* color-blind safe hue selection ensured the **C2**: *intuitive order* and **C5**: *CVD robustness* criterion. For sequential color schemes, two strategies were used to select seven colors from the sequential color schemes: first, using the default spacing provided by `cols4all` [154], and second, selecting seven colors from the full set of 14 using three custom algorithms that adapt to different criteria. Since the luminance varies depending on the structure of the color scheme, it was fixed at the mean value of 50 during all selection algorithms for the hue selections.

Algorithmic hue selection for sequential color schemes We propose and apply the following three selection algorithms that optimize hue selection for sequential color schemes. The algorithms share the objective of avoiding small perceptual distances within the color scheme in order to enhance the **C3: discriminative power** [118]. Each algorithm employs a different rule to select the color which is excluded in each step. The algorithm ideas built upon strategy of finding maximal spanning tree by eliminating minimal distances. We employ this principle to gain a set of hues with maximum distances as follows:

1. The first algorithm maximizes smallest distance among hues to ensure **C3: discriminative power**. The algorithm iteratively removes colors with smallest distance. It identifies pairs of hues with the smallest color distance. Then the distances of both colors to their respective outer neighbors are compared. The color with the smaller distance to its outer neighbor is then removed. This is done until seven hues with largest distances are identified.
2. The second algorithm maximizes average distance among hues in order to improve **C4: uniformity**. It considers the sum of distances to the neighboring colors for each color as elimination criterion. Since colors at the outer edges of the scheme have only one neighbor, the corresponding distance is counted twice. The color with the smallest sum of distances to its neighbors is then removed. This procedure is repeated until only seven hue selections remain. This algorithm is intended.
3. The third algorithm optimizes average minimum distance to improve hues to ensure **C3: discriminative power**. In this case, the color with the smallest sum of distances to all other colors is removed. First, the distances from each color to all other colors are summed. Then, the color with the smallest global distance is identified and removed from the set of hue selections. This procedure is repeated until seven hue selections remain. This algorithm represents a global variant of the second algorithm.

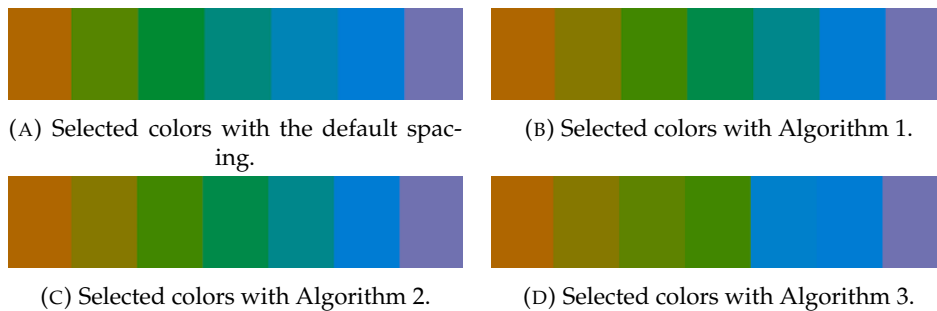


FIGURE 2: Example color schemes of the output of reduction Algorithms 1-3 for the viridis color scheme.

Selection of suitable hue schemes We define and apply elimination conditions (i.e, thresholds) based on the color-coding criteria **C1**, **C2** and **C4** based on the calculation of just noticeable differences (JNDs). JNDs describe the necessary difference between two colors, that can still be perceived by humans. Based on this, the elimination conditions are:

- **C1:** Legend-based order > 1 JND.
- **C1:** Legend-based order of colors under all three CVD conditions > 0.5 JND.

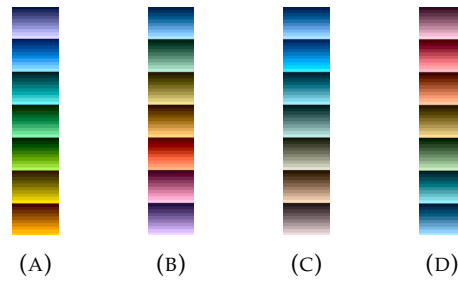


FIGURE 3: Hue schemes meeting the elimination criteria: (a) viridis, (b) batlow, (c) lapaz and (d) glasgow.

- **C2:** *Intuitive order by mantissa of each magnitude = 100%.*
- **C4:** *Uniformity within the magnitudes < 1 JND/step.*

Further we avoided schemes including a white, since the background for our meteorological data visualization is white.

Final selection The final selection of four hue schemes meeting the criteria is shown in [Figure 3](#).

3.2 Color Luminance Gradient Variation and Selection

We create a luminance gradient ranging from light to dark for each magnitude level. The luminance gradient forms the basis of color schemes for LVR, inspired by OMC. A monotonically changing luminance gradient from light to dark supports *high spatial frequency* [132] and ensures **C2: local intuitive order** [2]. This reinforces the *semantic association* through the dark-is-more bias [53, 141].

Our luminance gradient design recommendations follow the following **optimization criteria**:

- Limitation on luminance variation ensures that the brightest and darkest colors remain distinguishable, thus preserving **C1: legend-based order** and enhancing **C3: global discriminative power**
- A continuous luminance gradient across the entire color scheme is beneficial for supporting **C2: global intuitive order**.
- Strong luminance variation within individual magnitude levels increases **C3: local discriminative power**.
- Increased luminance variation the higher the magnitude levels contributes to **C4: uniformity**.

Design variations – proposals During the design process, various luminance gradient variation designs were explored to achieve these improvements. As an example, [Figure 4](#) shows the proposed variants of luminance gradients based on the color scheme batlow.

1. represents the original OMC color scheme design (denoted as "O" in the following). It has luminance uniformly reduced by 80 units within each magnitude level, ranging from 90 to 10. Luminance gradient variation. This version focuses on achieving **C3: discriminative power** and **C5: CVD robustness**

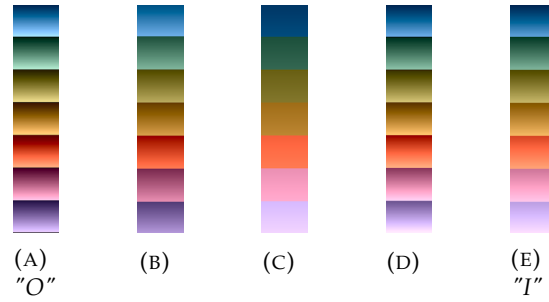


FIGURE 4: Examples of the designed luminance gradient variations defined by different luminance gradients, each applied to hue selections derived from the batlow color scheme: (a) constant luminance variation of 80 (denoted as Original-“O”), (b) constant luminance variation of 40, (c) decreasing luminance levels, (d) constant luminance variation & decreasing levels, (e) increasing luminance variation & decreasing levels (denoted as Increasing – “I”).

2. applies a constant luminance variation of 40 units within each magnitude level, spanning from 70 to 30.
3. decreases luminance continuously across the entire color scheme, resulting in a total range of 70 units (90 to 20) with a narrow local variance of 10 units.
4. implements a uniform luminance change of 60 units per magnitude level, with the lowest magnitude level ranging from 100 to 40 and each subsequent level decreasing by 5 units in both bounds.
5. varies both luminance level and range across magnitude levels, starting from 95 to 70 and gradually increasing to 65 to 10 in the highest magnitude. It focuses on achieving **C2: intuitive order** and **C4: uniformity**. This method is *denoted as “I”* below.

Resulting selection according to color-coding criteria The measurements corresponding to the criteria show that luminance gradient variation (a) best fulfills the criteria for *CVD safety* and **C3: discriminative power**, while structure (e) performs best in terms of **C2: intuitive order** and increasing **C4: uniformity**. Structure (c) excels in selected metrics but performs poorly in **C1: legend-based order** and **C3: discriminative power**, making it unsuitable. Structures (b) and (d) produce moderate results but are consistently outperformed by others.

As a result, two luminance gradient variations are selected for further analysis: the original OMC luminance gradient variation (a) and the increasing luminance gradient variation “I” (e). These represent complementary design focuses, offering a strong balance across conflicting criteria **C2, C3, C4 and C5**.

The full table of results for the elimination conditions of hues and luminance gradients can be found in the supplementary material.

3.3 Combination and Automatic Scheme Adaptation to Varying Number of Exponents

The final color schemes are produced by combining hue schemes and luminance gradient variations in Algorithm 1. Then, we propose another algorithm for adapting the color schemes to a required number of exponents (see Algorithm 2). This is an improvement over the fixed number of exponents used in the OMC color scheme [21].

3.3.1 Color Scheme Creation

To create an effective LVR color scheme from the improved hue selections and luminance gradient variations, Algorithm 1 is proposed. The selection of hues to be chosen from includes viridis (V), batlow (B), glasgow (G), and lapaz (L), in combination with the original (O) and increasing (I) luminance gradient variations (the resulting color values can be found in the supplementary material).

Algorithm 1 Algorithm for creating the novel LVR color scheme.

Input: Hue and chroma values, luminance range per magnitude

Output: Array containing hue, chroma and luminance for all points of the color scheme

Add each hue value to array hues twice;

Add each chroma value to array chromas twice;

Add the highest and lowest luminances value per magnitude to array luminance;

$n = \text{length}(\text{control_points});$

$\text{positions} = [0, 99, 100, 199, \dots, 100 \times n, 100 \times n + 99];$

$\text{steps} = \max(\text{positions}) + 1;$

$\text{interpolated_hues} = \text{linear_interpolation}(\text{luminances}, \text{positions}, \text{steps});$

$\text{interpolated_chromas} = \text{linear_interpolation}(\text{luminances}, \text{positions}, \text{steps});$

$\text{interpolated_luminances} = \text{linear_interpolation}(\text{luminances}, \text{positions}, \text{steps});$

Combine interpolated_hues, interpolated_chromas,

interpolated_luminances

into array color_scheme;

Return color_scheme;

Normalization and Denormalization Before applying our novel color scheme to data, it must first be logarithmized through a normalization and denormalization process. This is necessary to correctly map the values onto the intervals because the color scheme maps magnitudes to linear intervals between the mantissa steps rather than using standard logarithmic scaling. This makes the color differences between the mantissa steps more uniform and easier to read. Therefore, Equation 1 and Equation 2 were implemented. For the normalization, the exponent is first computed, and the corresponding mantissa value is proportionally scaled relative to the nine distinct mantissa levels. Normalization of numeric values:

$$\begin{cases} \log_values, & \text{if mantissas} = 1; \\ \text{exponents} + (\text{mantissas} - 1)/9, & \text{otherwise;} \end{cases} \quad (1)$$

After assigning color values, the exponent and mantissa are extracted from the normalized value through the denormalization process and the original input is reconstructed as $\text{mantissa} \cdot 10^{\text{exponent}}$. Values with a mantissa equal to 1 require special treatment to avoid division by zero.

Denormalization of normalized values:

$$\begin{aligned} \text{mantissas} &= \begin{cases} 1, & \text{if fractions} = 0; \\ 1 + 9 \cdot \text{fractions}, & \text{otherwise;} \end{cases} \\ \text{values} &= \text{mantissas} \cdot 10^{\text{exponents}} \end{aligned} \quad (2)$$

3.3.2 Adaption to Lower Number of Exponents:

Since the current selection of hues and the definitions of the luminance gradient variations are based on seven orders of magnitude, these components must be adjusted when fewer magnitude levels are required.

To achieve this, hues are iteratively removed from the original set of seven hues, always eliminating the one with the smallest perceptual distance to its immediate neighbors as described in Algorithm 2.

Algorithm 2 Algorithm for a reduced number of magnitudes with the original luminance gradient.

```

Input: 7 hues with luminance = 50, number of magnitudes
Output: Reduced set of hues with size = number of magnitudes
while  $length(hues) > num\_magnitudes$  do
    Compute the CIEDE2000 distance matrix
    Identify the pair of adjacent colors with the smallest distance
    if the pair includes a color at index 0 then
        Remove the color at index 1 from hues
    else
        The pair includes a color at index  $length(hues)$ 
        Remove the color at index  $length(hues) - 1$  from hues
    else
        Compare the distance of each color in the pair to its outer neighbor
        Remove the color with the smaller external distance from hues;
    end if
end while
return hues

```

4. Evaluation

We evaluate the novel color scheme options (see Figure 6) in two ways: 1) Quantitatively based on measurable color-coding criteria and 2) empirically by conducting a human-subject study. Both ways compare the novel schemes with the original OMC color scheme by Braun [21].

4.1 Quantitative Evaluation

We calculate and compare the measurable criteria **C1-C5** for each novel color scheme and the baseline original OMC color scheme. Criterion **C6**, as well as the aesthetic aspect, cannot be measured and is therefore tested in a human subject study below.

To ensure comparability, the value range adaptation is applied to the novel color schemes in order to reduce them to six orders of magnitude, matching the range covered by the original OMC color scheme.

Method: The calculation measures the quality of the proposed color schemes with respect to the color-coding criteria **C1-C5**. For this purpose, we reformulated each color-coding criterion into a quantitative measure, as described in Section 3.1. Then, we calculated these values for each combination of hue scheme and luminance gradient. Results are shown in Figure 5. It highlights criteria that are met in green color, and the two best-performing combinations are outlined in dark green.

Color Scheme Variations/Color-Coding Criterion		C1: Legend-Based Order	C2: Intuitive Order				C3: Discriminative Power		C4: Uniformity		C5: Color Vision Deficiency Robustness			
Hue Selection	Luminance Gradient Variation	Global Legend-Based Order	Global Intuitive Order	Local Intuitive Order	Global Intuitive Order by Mantissa	Local Intuitive Order by Mantissa	Global Discriminative Power	Local Discriminative Power	Global Uniformity in Magnitudes	Increasing Uniformity	CVD Robustness Average	CVD Robustness Deuteranomaly	CVD Robustness Protanomaly	CVD Robustness Tritanomaly
Original	Original	1.82 JNDs	47.31 %	84.48 %	45.0 %	75.0 %	2.80 JNDs/step	8.07 JNDs/step	1.05 JNDs/step	False	1.12 JNDs	0.68 JNDs	0.94 JNDs	1.75 JNDs
Viridis	Original	5.26 JNDs	58.46 %	82.76 %	75.0 %	100.0 %	3.81 JNDs/step	12.41 JNDs/step	0.82 JNDs/step	False	1.01 JNDs	0.49 JNDs	0.99 JNDs	1.56 JNDs
Viridis	Increasing	1.53 JNDs	88.21 %	82.76 %	90.0 %	100.0 %	2.56 JNDs/step	5.74 JNDs/step	0.31 JNDs/step	True	1.06 JNDs	1.43 JNDs	0.24 JNDs	1.52 JNDs
Batlow	Original	5.26 JNDs	44.83 %	82.76 %	40.0 %	100.0 %	3.81 JNDs/step	12.49 JNDs/step	0.82 JNDs/step	False	1.56 JNDs	0.14 JNDs	1.21 JNDs	3.33 JNDs
Batlow	Increasing	1.53 JNDs	82.58 %	82.76 %	95.0 %	100.0 %	2.62 JNDs/step	5.92 JNDs/step	0.31 JNDs/step	True	1.05 JNDs	0.10 JNDs	1.44 JNDs	1.62 JNDs
Glasgow	Original	5.26 JNDs	51.45 %	82.76 %	60.0 %	100.0 %	3.63 JNDs/step	12.38 JNDs/step	0.82 JNDs/step	False	1.35 JNDs	0.59 JNDs	1.04 JNDs	2.40 JNDs
Glasgow	Increasing	1.53 JNDs	87.45 %	82.76 %	95.0 %	100.0 %	2.44 JNDs/step	5.71 JNDs/step	0.31 JNDs/step	True	1.24 JNDs	0.65 JNDs	1.56 JNDs	1.50 JNDs
Lapaz	Original	5.26 JNDs	47.64 %	82.76 %	55.0 %	75.0 %	3.33 JNDs/step	12.24 JNDs/step	0.82 JNDs/step	False	0.98 JNDs	0.67 JNDs	0.89 JNDs	1.38 JNDs
Lapaz	Increasing	1.53 JNDs	85.23 %	82.76 %	100.0 %	100.0 %	2.07 JNDs/step	5.41 JNDs/step	0.31 JNDs/step	True	1.17 JNDs	1.51 JNDs	0.44 JNDs	1.56 JNDs

FIGURE 5: Quantitative analysis of the eight best criteria-based color schemes and the original OMC color scheme. All measurements were computed with six orders of magnitude and all combinations of hue selections and luminance gradient variations are included.

Results: Figure 5 shows the main results. Improvements achieved by the novel color schemes in comparison to the original OMC color scheme can be observed for the following criteria:

- **C1: Legend-based order:** Improved by all novel color schemes with the original luminance gradient variation.
- **C2: Global intuitive order:** Improved by all novel color schemes except for the color scheme using batlow hue selections combined with the original luminance gradient variation (B0).
C2: Global and local intuitive order by mantissa: Improved by all novel color schemes except for the color scheme using lapaz hue selections with the original luminance gradient variation (L0).
- **C3: Global and local discriminative power:** Improved by all novel color schemes with the original luminance gradient variation.
- **C4: local uniformity within magnitudes:** Improved by all novel color schemes.
C4: global increasing uniformity: Improved by all novel color schemes using the increasing luminance gradient variation.
- **C5** The best result is achieved by the batlow hue selections with the original luminance gradient variation (B0). all schemes but LO meet the threshold of an average **C5: CVD robustness** of at least one JND.

Recommended Color Schemes Both glasgow-based color schemes (Glasgow-Original (G0) and Glasgow-Increasing (GI)) consistently perform better or at least similarly to the original OMC color scheme, depending on the luminance gradient variation used. Thus, they are recommended.

4.2 Human-Subject Evaluation

We evaluate the novel color schemes in Figure 6 across various visualization types in a human-subject study. To achieve comparability, all color schemes were tested with six orders of magnitude. We measure error rates and response time as well as aesthetic rankings and color name variations. Accordingly, we test for the following **hypotheses**:

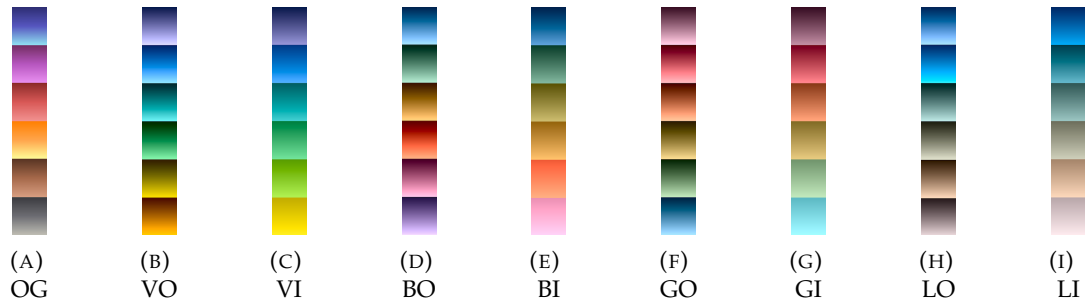


FIGURE 6: All color scheme options tested in the human-subject study (plus the original OMC color scheme OG), collected by all hue adaptations and luminance gradient combinations. Hue options include V: viridis, B: batlow, G: glasgow, L: lapaz and luminance gradients include "O": original luminance gradient variation, "I": increasing luminance gradient variation.

- **H1:** The novel color schemes result in *fewer errors* than the original OMC color scheme in identification, localization, and comparison tasks.
- **H2:** The novel color schemes *reduce the time* required to solve identification, localization, and comparison tasks compared to the original OMC color scheme.
- **H3:** The novel color schemes receive *better aesthetic rankings* than the original OMC color scheme.
- **H4:** The novel color schemes exhibit a higher **C6: color name variation** than the original OMC color scheme.

4.2.1 Study Design

Data We used synthetic data as it allowed for control over the distribution across the orders of magnitude. For example, data with higher values (and darker colors) could potentially be more difficult to identify. A higher prevalence of large values could thus negatively affect the performance and lead to misinterpretations of the user study results. This concern can be excluded since all orders of magnitude were evenly distributed in the visualizations.

A large value range of 1 – 1000000 was chosen. This creates six orders of magnitude with exponents ranging from zero to five. The distances of the values were selected so that all integer mantissa values within each order of magnitude should be correctly identifiable. The visualizations each display all integer mantissa values in every exponent (except the choropleth visualizations, in which the number of data values is fixed). This results in 54 values that are randomly distributed in the visualizations to generate different visual representations and prevent inferences about the values based on the arrangement.

Stimuli Design: Color Schemes We tested all nine different previously described color schemes (see Figure 6). All stimuli use the same color scheme design regarding labels and ticks to ensure that differences in performance cannot be attributed to variations in labeling or other unrelated factors. For all color schemes we used a linear scaling within each order of magnitude for a more intuitive identification of mantissa values. As in the original OMC design [21], all labels are written in scientific notation. The scheme is labeled with the generic term *value*, in order to avoid introducing any semantic associations that might influence interpretation and performance [2].

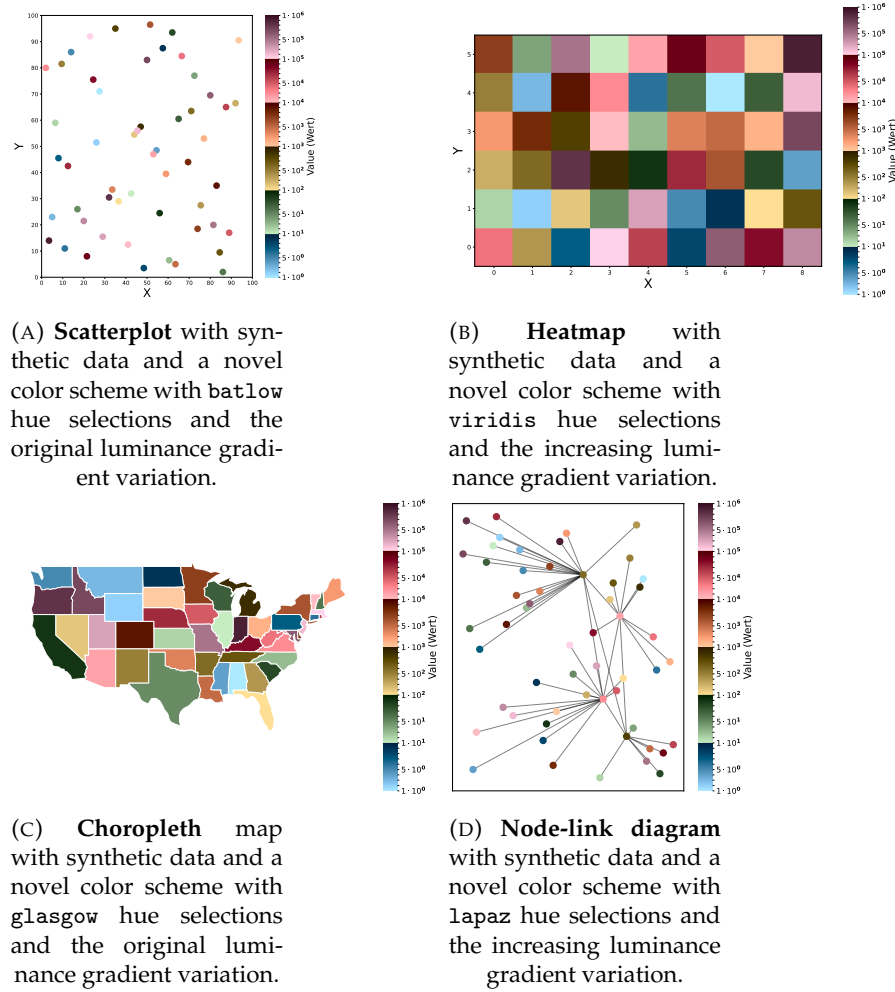


FIGURE 7: Examples of all four visualization types used in the human-subject study with synthetic data and novel color schemes.

Stimuli Design: Visualization Types The color schemes were tested on multiple visualization types rather than just (meteorological) scatterplots by Braun et al [21]. This approach tests the applicability novel schemes to more general use cases and a broader range of visualization types. The study incorporates four distinct **visualization types** that commonly utilize color-coding for quantitative values (see Figure 7).

Scatterplot: Scatterplots represent the baseline visualization form, which was also used in the original OMC evaluation [21], making them essential for direct comparison with previous findings. The scatterplots were created with both the x- and y-axis spanning from zero to 100. Both axes were labeled with the generic terms X and Y to avoid introducing misleading or distracting titles. Each of the 54 data points was assigned to a random x and y position, each within the range of two to 98. A minimum distance of 1.5 units between two points was enforced. These constraints ensure that the values do not overlap excessively or appear too close to the plot boundaries, which could result in incomplete rendering.

Heatmap: Heatmaps were selected as they are essential for the analysis of scientific data and can only be effectively interpreted with appropriate color-coding [58]. The heatmap uses x-values ranging from zero to eight and y-values ranging from zero to five, providing a total of 54 unique positions. As with the scatterplot, only the axes were labeled with

the generic terms X and Y . Each value was randomly assigned to one of these positions.

Choropleth: Choropleth maps were included as they represent a central application domain for color-coding [57]. The choropleth map visualizations of the United States include 49 different states [40]. Textual labels were intentionally omitted in order to keep the visualization as general and neutral as possible, and to avoid introducing any potential bias or misinterpretation. Since it is not feasible to assign all 54 values to a state or an equal number of values per order of magnitude, five values out of the 54 values are randomly chosen and excluded. All remaining values are randomly assigned to one of the states.

Node-link diagram: Node-link diagrams offer an effective way to visualize the relationships within scientific data and were included due to their potential for enhancement through color-coding [134]. For the diagram generation, scatterplots serve as the basis for determining both the positions and color-coding of the nodes as described above. Edges were generated randomly according to the following procedure: Six central nodes were selected and connected to all closest surrounding nodes. To avoid excessive distances between the nodes and their assigned centers, the central nodes were evenly distributed within the visualization. This was achieved by choosing the nodes closest to the evenly spaced positions [35, 25], [35, 50], [35, 75], [65, 25], [65, 50], and [65, 75] as the six centers. Each center was also connected to the nearest other center.

Tasks We tested six tasks that were selected focusing on basic task types that allow for evaluating fundamental color-coding applicable in a general context [156]. Moreover, they are common in large value ranges and color-coding research [16, 19–21, 65]. Four readability tasks had to be solved using the visualizations described above:

- **Identification:** Participants were asked to identify the mantissa and exponent values of a randomly chosen data point that was marked with an arrow. (See Figure 8 for the input interface)
- **Localization:** Participants had to identify the data points corresponding to the highest and lowest values within a marked area containing four values, and provide their mantissa and exponent values. Two levels of difficulty are contained in this task: 1) Values in the marked area all within the same order of magnitude, 2) Values in the marked area all from different orders of magnitude.
- **Exponent comparison:** Participants had to select which of two marked values (from different orders of magnitude) is higher.
- **Mantissa comparison:** Participants had to select which of two marked values (both in the same orders of magnitude) is higher.

Please enter the **value of the marked datapoint**.



Mantissa · 10 Exponent

FIGURE 8: Participants had to manually enter values for the mantissa and exponent for certain points marked in the visualization.

Two additional tasks evaluated non-measurable color scheme quality through subjective quality reviews:

- **Aesthetics:** Participants were asked to rank the color schemes according to their aesthetic appeal.
- **Color name variation (C6):** Participants were asked to choose a name for each hue selection of the color schemes from a single-choice multiple-options list of a predefined set of color terms. The total of 11 terms were taken from the list identified by Kim et al. [81]. The hue selections were evaluated with a fixed luminance level of 50. As a result, only five color schemes were tested instead of nine, since the novel color schemes include each set of hue selections twice with both luminance gradient variations.

4.2.2 Experimental Setting and Participants

The study was conducted online. For a reduced study completion time per participants, we opted for a **between-subject study design**, with the groups defined by the different visualization designs. Therefore, each participant saw each color scheme and had to complete 42 trials in total. To prevent learning effects, the order of the stimuli were randomized with respect to the color schemes [30]. Example questions from the study can be found in the supplementary material.

The **study procedure** began with demographic questions and a question about participants' experience with relevant technical concepts (data with large value ranges, scientific visualizations, logarithmic scaling, and color theory). Moreover, to ensure that the devices used to conduct the user study were capable of displaying color gradations appropriately [147], a simplified version of the Eizo test [51] was used. The test consisted of an image showing eight distinct color steps. Participants were asked to report how many of these steps they were able to perceive. To test for color vision deficiencies, we used Ishihara and Optical Hardy-Rand-Rittler plates [103]. The actual study trials started after an initial training phase on scientific notation, the logarithmic color-coding, and the visualization type of the respective between-subject group. In between the trials, two attention checks were included. For each trial, we measured the participants' response time.

Participants were recruited through advertisements in lectures and did the study voluntarily. After applying all exclusion criteria, the responses of 76 participants remained for evaluation. The distribution of participants across the different visualization types, corresponding to the randomization groups, was approximately balanced (18 scatterplot, 19 choropleth, 20 heatmap, and 19 node-link). A total of 46 participants completed the user study on a phone, while 30 participants used a laptop. The gender distribution was almost balanced (F: 54%, M: 43%, no answer: 3%) and the majority were between 20 and 40 years old (71%). Their educational levels ranged from high school diplomas to masters degrees. While 57% of the participants had no experience with underlying concepts of the study topic, 22% of the participants had experience with the four concepts (named above) on average.

4.2.3 Analysis

Error Measurement In the *identification and localization* tasks, participants were asked to enter mantissa and exponent values. Subsequently, we applied the quantitative error metric introduced by Braun et al. [21]. With $v_{\max}$ and $v_{\min}$ indicating the higher and lower value from the given and correct answer and $e(v)$ and $m(v)$ denoting exponent

and mantissa of a value, the error is computed as follows:

$$\text{error} = e(v_{\max}) - e(v_{\min}) + \frac{m(v_{\max}) - m(v_{\min})}{9} \quad (3)$$

This approach is well suited for large value ranges, as it treats deviations in exponent and mantissa uniformly across different orders of magnitude. Errors smaller than 1 represent mantissa errors, integer values greater than or equal to one indicate errors in the exponent and real-valued errors equal to or greater than one indicate an error in both exponent and mantissa.

In the localization task, participants were required to report both the lowest and highest value within a defined range. Here, the error was calculated separately for each of the two responses using Equation 3, and the mean of both values was used as the final error measure.

For the *comparison* tasks, participants selected whether the marked data point *a* had the higher value, data point *b* had the higher value, or both values were equal. As responses in these tasks can be classified strictly as correct or incorrect, the error was computed as the percentage of incorrect answers.

Significance Tests All statistical tests were conducted with the standard significance level of $\alpha = 0.05$. First, all quantitative datasets were tested for a normal distribution using the Shapiro–Wilk test. The results concluded that none of the datasets follow a normal distribution. Therefore, we opted for non-parametric hypotheses tests.

For *quantitative and ordinal datasets*, we ran Kruskal–Wallis tests for a general presence of significant differences and Mann–Whitney–U tests as post-hoc tests for identifying pairwise differences. For *categorical datasets*, Chi-squared tests were conducted both for significant main effects and post-hoc pairwise comparisons. Bonferroni correction was applied to all pairwise tests.

4.2.4 Results and Discussion for Error Rates

For each task, we analyze error (inaccuracy) and response time, and provide a ranking of the proposed designs based on the results. In the summaries of the pairwise significant differences, the symbol $\prec$ denotes a significantly worse performance of the color scheme on the left compared to the color scheme on the right.

Figure 9 shows the errors rates for the identification (a), localization (b), exponent comparison (c), and mantissa comparison (d) tasks.

Identification task: V0 and B0 achieve the lowest error with a median of 0, whereas all other color schemes show a slightly higher median corresponding to one mantissa step. Regarding the inter quartile range, V0 and B0 also show the most compact distribution, indicating lower variability in the responses. The Kruskal–Wallis test indicated a statistically significant difference across the color schemes ($p < 0.01$). A detailed table of the results of the pairwise comparisons using Mann–Whitney–U tests are presented in the supplementary material. The pairwise statistically significant differences are:

- BI $\prec$ V0, B0
- LI $\prec$ V0, VI, B0, G0
- OG $\prec$ V0, B0

In sum: V0, B0 outperform the other tested color schemes.

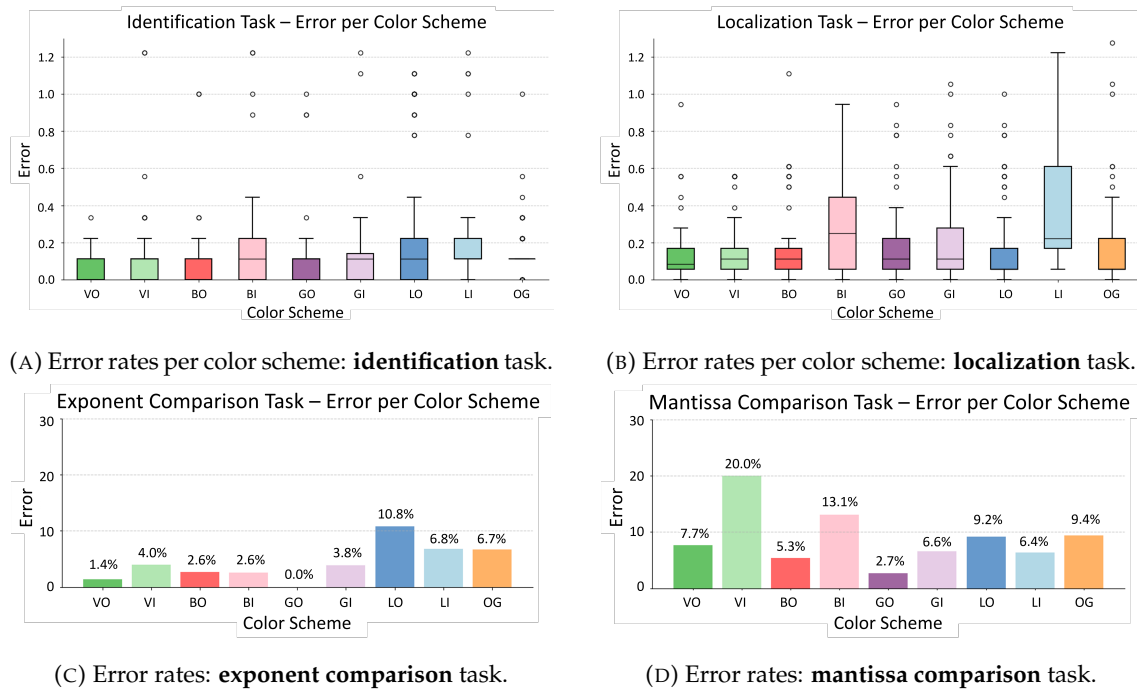


FIGURE 9: **Error rates** of the identification (a), localization (b), exponent comparison (c), and mantissa comparison (d) task per color scheme (*the lower, the better*).

Localization task: The color schemes L0 and OG achieve the lowest median error, approximately 0.056 (see Figure 9(b)). V0 follows with a slightly higher median error of approximately 0.083. BI and LI lead to the highest error rates greater than 0.2. The color scheme L0 exhibits lower variance than OG, with values more tightly clustered around the median. The Kruskal–Wallis test indicated a statistically significant difference across the color schemes ($p < 0.01$). The results of the pairwise comparisons using Mann–Whitney–U tests are presented in the supplementary material. These statistically significant differences in performance for the localization task can be summarized as follows:

- BI $\prec$ V0, L0, OG
- LI $\prec$ VI, V0, BO, GI, GO, L0, OG

In sum: V0, OG and L0 outperform the other tested color schemes.

Exponent comparison Task: The lowest error rate is achieved by G0 with 0% incorrect responses (see Figure 9(b)). This is followed by V0 with 1.4% incorrect answers, and BI with 2.6%. The highest error rates has L0 (10.8%). A chi-squared test yields a test statistic of $\chi^2 = 15.412$ and a p -value of 0.052, which is slightly above the predefined significance level of $\alpha = 0.05$.

In sum, no statistically significant differences were found between the color schemes.

Mantissa comparison task: The color scheme G0 achieves the lowest error rate with only 2.7% incorrect responses, followed by B0 with 5.3% errors (see Figure 9(d)). VI and BI yielded the highest error rates.

The chi-squared test detected a significant mean effect between the color schemes ($p = 0.016$). However, the results of the pairwise chi-squared post-hoc tests revealed *no statistically significant differences* between pairs of color schemes due to the Bonferroni correction.

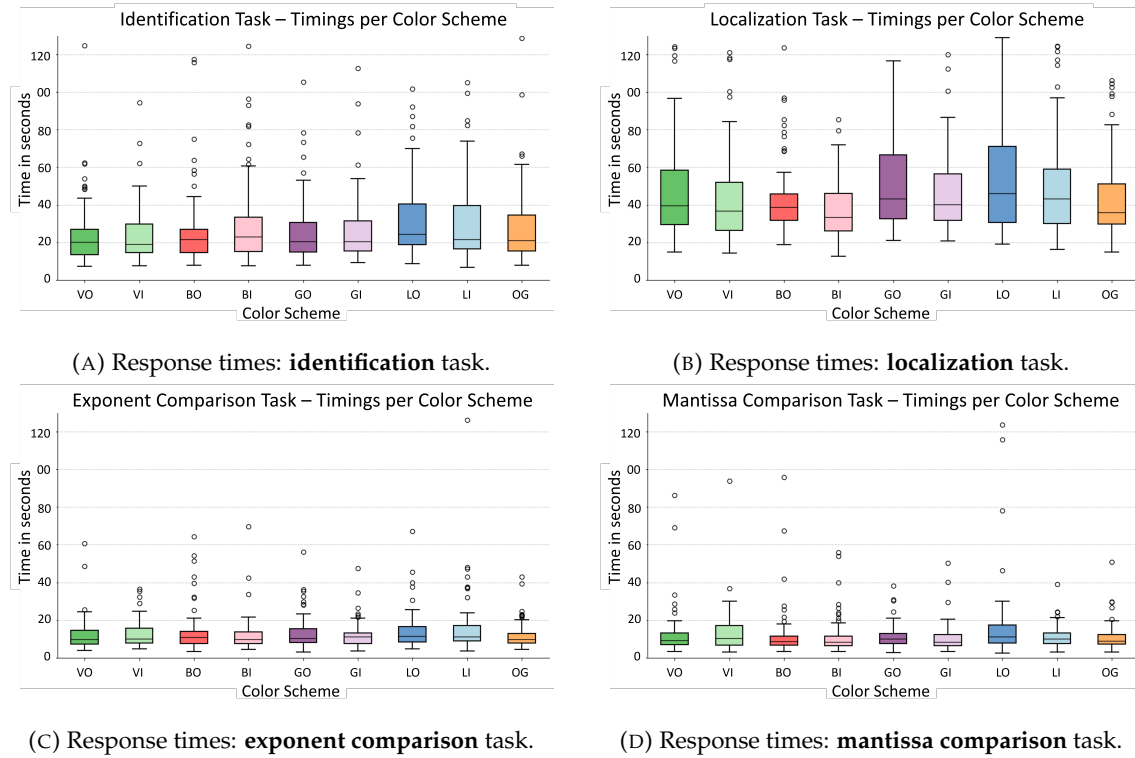


FIGURE 10: **Response times** for solving the identification (a), localization (b), exponent comparison (c), and mantissa comparison (d) task per color scheme (*the lower, the better*).

Discussion and Hypothesis Results Following above-presented results, **hypothesis 1** can be partly confirmed for the identification task, where the color schemes VO and BO achieved significantly better results than the original color scheme OG. The remaining color schemes performed equally well, without significant differences.

For the localization task, the color schemes BI and LI performed significantly worse, while all other color schemes showed no significant difference compared to the original OMC color scheme OG. Therefore, the hypothesis cannot be confirmed for this task at all. These findings are consistent with the results reported by Gołbiowska et al. [53], who showed that color schemes adapted according to the **C3: discriminative power** are more suitable for reading tasks, while color schemes with a focus on **C2: intuitive order** are better suited for recognizing shapes. With regard to our evaluated options, a focus on **C2: intuitive order** was observed in the increasing color schemes, whereas a focus on **C3: discriminative power** was associated with the original luminance gradient variation. Accordingly, significantly better performance was observed with the original luminance gradient variation and significantly worse performance with the increasing luminance gradient variation in the reading tasks identification and localization. For the exponent comparison and mantissa comparison tasks, no statistically significant differences in performance were observed, and the hypothesis must also be rejected for these tasks.

4.2.5 Results and Discussion for Response Times

For each task, we analyze the response time, and provide a ranking of the proposed designs based on the results. In the summaries of the pairwise significant differences,

the symbol $\prec$ denotes a significantly worse performance of the color scheme on the left compared to the color scheme on the right.

Figure 10 presents the time required to complete the four tasks. Times are in seconds, and the visualizations display values up to 120s for readability, leaving out outliers that are higher than this threshold.

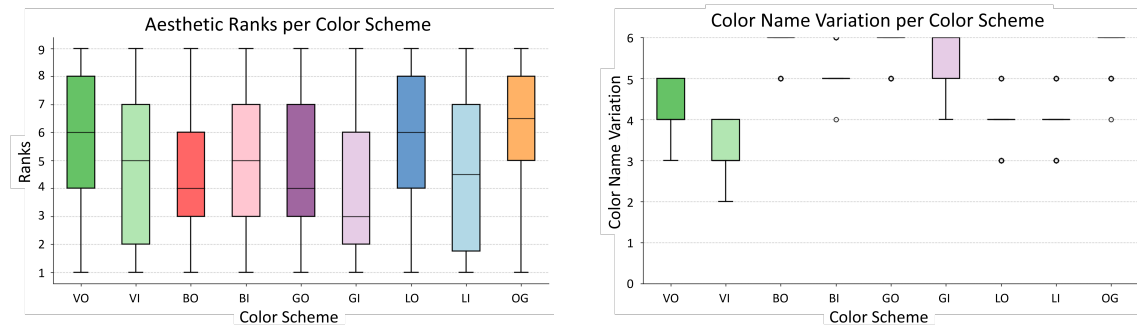
For the **identification and the comparison tasks**, the median values of all color schemes lie within narrow ranges. For *identification*, VI lead to the lowest median response time of about 19 seconds, while L0 had the highest response time of about 24 seconds. The comparison tasks show low timing intervals, with medians ranging from 9.7 to 11.4 seconds for the *exponent*, and from 8.4 to 11.2 seconds for the *mantissa* comparison. The Kruskal–Wallis test indicated *no significant difference in response times* for these three tasks ($p_{ident} = 1.0$, $p_{exp} = 0.54$, $p_{mant} = 0.11$). The assumption that the mantissa comparison task is more difficult than the exponent comparison task. It would, thus, require more processing time, is not supported by the timing results.

For the **localization task**, the response times have a higher variance between the color schemes. The medians range from 33.42 seconds with BI to just under 40 seconds with L0. This overall increase of response times can be attributed to the task requiring participants to report two values, each with mantissa and exponent, instead of just one. The Kruskal–Wallis test showed a significant mean effect ($p < 0.01$). The pairwise comparison lead to the following statistically significant differences in terms of required time for the localization task:

- $\{G0, L0\} \prec BI$

Discussion and Hypothesis Results Hypothesis 2 assumes that the novel criteria-based color schemes reduce response times compared to the original OMC color scheme. Across all tasks involving timing measurements, no statistically significant differences between the improved and the original OMC color schemes were found. Accordingly, the hypothesis cannot be confirmed. A similar observation was made in the user study on the original OMC color scheme [21].

4.2.6 Results and Discussion for Aesthetics Rankings



(A) Aesthetic ranks (the lower, the better).

(B) C6: Color name variation (the higher, the better).

FIGURE 11: Participant rankings on *Aesthetics* and *Color name variation* per color scheme.

The aesthetic rankings as evaluated by participants during the user study are shown in Figure 11a. Lower rank values indicate better aesthetic perception. As nine color schemes were compared, possible ranks range from 1 (most aesthetic) to 9 (least aesthetic).

The color scheme GI achieves the best result with a median rank of 3. B0 and G0 follow with median ranks of 4, while V0 and OG are perceived the least appealing with median ranks above 6.

The Kruskal–Wallis test confirmed the presence of statistically significant differences in aesthetic preference ($p < 0.01$). The results of the pairwise Mann–Whitney–U post-hoc tests show significantly better aesthetic outcomes for novel color schemes compared to the original OMC color scheme OG:

- $V0 \prec B0$
- $V0, L0 \prec GI$
- $OG \prec VI, B0, G0, GI, LI$

Hypothesis 3 focuses on aesthetic perception and states that the novel criteria-based color schemes are rated significantly more aesthetically pleasing than the original OMC color scheme. This can be confirmed for the color schemes VI, B0, G0, GI, and LI, which all received significantly better aesthetic rankings than OG. No significant differences were observed for the remaining color schemes.

4.2.7 Results and Discussion for Color Name Variation

The color name variation can range from a minimum value of 1, indicating that all hue selections were assigned the same name, to a maximum of 6, where each hue selection received a distinct name. The results, expressed as the number of distinct color names assigned per color scheme, are shown in [Figure 11b](#).

The color schemes B0, G0, GI, and OG achieved the highest possible **C6: color name variation** in the median and thus represent the best performing color schemes for this criterion. Among them, B0, G0, and OG exhibited no variance in the distribution, aside from individual outliers, indicating a consistently high level of perceptual distinction.

The Kruskal–Wallis test indicated a significant main effect in the **C6: color name variation** ($p < 0.01$). The pairwise results of the post-hoc Mann–Whitney–U test show the following significant differences between color schemes:

- $VI \prec L0, LI \prec V0 \prec BI \prec GI \prec G0, OG, B0$

Hypothesis 4 states that the **C6: color name variation** is higher for the novel criteria-based color schemes than for the original OMC color scheme. The performances can be ranked in ascending order, with G0, OG and B0 achieving the maximum possible **C6: color name variation** of six distinct names. All other color schemes showed significantly lower variation. Thus, this hypothesis is rejected.

4.3 Summary of Results

All in all, the color schemes B0 and G0 consistently achieved results that were either equivalent to or significantly better than those of the original OMC color scheme across all tested hypotheses. They thus represent superior alternatives to the original OMC color scheme according to the hypotheses tests. This confirms that empirically measurable improvements can be made by the novel color schemes B0 and G0. The quantitative analysis shows the same pattern in favor of G0, meaning it achieved either equivalent to or significantly better performances than with the original OMC color scheme. As a result, the use of the G0 color scheme, consisting of glasgow-based hue selections and the original luminance gradient variation, can be recommended before the original OMC color scheme according to the evaluations.

All color schemes based on glasgow hue selections consistently performed at least as well as the original OMC color scheme and, in several aspects, demonstrated measurable improvements: Both of our luminance gradient variations achieved enhancements in terms of **C4.1: global uniformity** and **C2: intuitive order** when combined with the glasgow hue selections. The improvements in **C4.1: global uniformity** are due to the continuous luminance gradient being applied across the color scheme, since it leads to uniform changes in the colors [89]. The improvements in **C2: intuitive order** observed in both luminance gradient variations can be attributed to the pre-sorting of hues.

The increasing luminance gradient variation further improved **C2: intuitive order** by incorporating overall higher luminance levels for higher magnitudes [2, 174]. By applying the increasing luminance gradient variation, the **C4.1: global uniformity** could also be fulfilled, which is caused by stronger luminance variations in higher orders of magnitudes. Higher luminance variations in magnitudes with a higher value range lead to more changes in the colors which represent the underlying changes in the data better [89, 109].

The original luminance gradient variation led to additional improvements in **C1: legend-based order** and **C3: discriminative power**. These effects are a direct result of selecting hue selections with large perceptual distances [109, 157], realized through choosing color schemes in the `cols4all` tool with a broad hue spectrum. Moreover, the high luminance variation in each magnitude, defined by the original luminance gradient variation, increases color distances and therefore the **C1: legend-based order** and **C3: discriminative power** [118].

5. Limitations and Future Work

We investigated criteria-based improvements for the color-coding of large value ranges. Our approach was evaluated both quantitatively and through a human-subject study. Our generalizable results, together with identified limitations, suggest several directions for future research.

Color scheme creation: To generate and improve the color schemes, we used general color-coding criteria and applied them to the specific characteristics of large value ranges. In order to keep the color nameability and distinguishability of the hue selections as high as possible [126], the measurements were computed with color schemes containing the maximum of seven orders of magnitude. For comparability with the original OMC color scheme, the quantitative analysis was conducted with six orders of magnitude. However, performance may vary for color schemes with fewer or more magnitudes.

Our work focused only on unipolar data. Thus, the resulting color schemes have limited applicability to bipolar data, since no perceptual midpoint is highlighted and relevant information may be lost [57]. An extension of our approach to bipolar data would add applicability of the color schemes to datasets with large value ranges that span both positive and negative exponents.

Study design: We conducted our human-subject study employing common visualization designs and common readability tasks to ensure generalizability of the results. All data used in the user study was synthetically generated with controlled distributions for a balanced representation of magnitudes. While this was necessary for a systematic evaluation, such distributions may not reflect various real-world dataset, which could be more irregular. The complexity of the tasks had to be limited to maintain a manageable study duration, considering the large number of color schemes compared. In future, the color schemes could be tested for more complex or high-level tasks.

As the study was conducted online, the viewing environment could not be fully controlled. Device characteristics [143, 147, 156], ambient lighting conditions [174], and

screen backgrounds [141] might influence color perception and, consequently, performance. Although visual diagnostics such as the Eizo monitor test [51] were employed to increase control, device-related variability remained, and full standardization could not be achieved. Repeating the study in a laboratory setting would overcome this problem.

Given the important role of color vision deficiencies in the accessibility of color-coding, they were incorporated as a criterion in the design process (C5: *CVD robustness*) and evaluated in the quantitative analysis. Due to the small number of participants with a specific type of CVD, this group could not be evaluated separately in the human-subject study. As a result, this aspect could only be addressed in the quantitative analysis, and not validated through empirical user data.

6. Conclusion

We presented an improved and automated design of effective color schemes for data with large value ranges. We ensured meeting a set of criteria-based design principles. Key improvements include the systematic selection of hues and refined luminance gradients for each magnitude. We evaluated multiple novel color schemes through a quantitative analysis of measurable criteria and a user study that focused on task performance, aesthetic perception, and color name variation.

The results of these evaluations identified Glasgow-original (G0) to be an effective color scheme that significantly outperforms the original OMC color scheme in specific criteria such as color discriminability and aesthetic appeal. For all remaining criteria, the performance is equally strong across the tested basic task types. At the same time the luminance gradient variation Glasgow-increasing (GI) color scheme significantly improves the uniformity of color distances and the perceptual order. A notable contribution of our work is the automation algorithm, which allows for the dynamic adjustment of the number of supported orders of magnitude within an extended range from two to seven exponents.

The novel color schemes and the automated adaptation algorithm provide robust methods for enhancing the clarity and effectiveness of visualizations in various applications.

7. Acknowledgments

This work was partially supported by the BMBF WarmWorld, SANE and Risk-Principe Projects. We thanks participants for taking part in the study and S. Bremm and D. Archambault for their constructive suggestions on the paper draft.

Chapter 3

Human Perspectives on the Use of Color in Data Visualization for Qualitative Data

3.1 Discriminability of Node Colors: Semantic Color Associations

3.1.1 Introduction

The following paper investigates various human strategies for creating color palettes for qualitative multi-label data with strong semantic color associations. Satisfying the need for semantic color assignments for multiple labels at the same time poses a particular challenge for the discriminability of nodes and point colors.

3.1.2 Human Strategies for Semantic-Based Colors Representing Multi-Label Classes

This paper [116] will be published and presented at the ACM Symposium on Applied Perception 2026.

Laura Pelchmann, Lena Jacobi, Sebastian Bremm, and Tatiana von Landesberger. 2026. Human Strategies for Semantic-Based Colors Representing Multi-Label Classes. In *ACM Symposium on Applied Perception 2026 (SAP'26)*, August 26–27, 2026, Rennes, France. ACM, New York, NY, USA, 10 pages. doi: [10.1145/3821409.3821477](https://doi.org/10.1145/3821409.3821477)

I am the primary author of this publication. In this role, I was responsible for the design, data analysis, as well as the writing and publication of the work. The individual contributions of myself and my co-authors to this publication are outlined below according to the Contributor Roles Taxonomy (CRediT):

Laura Pelchmann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft.

Lena Jacobi: Lena conducted her master's thesis on semantic color schemes, guided by me and supervised by Tatiana. In this regard, her research on semantic color associations on farm animals and the corresponding results are part of the following paper. Moreover, she proofread the final paper submission. Contributor Roles: Data curation, Formal analysis, Investigation, Software, Visualization, Writing – review & editing.

Sebastian Bremm: Sebastian has made important contributions through his expertise in

statistical analysis and study design. Moreover, he was always available to provide feedback and answer questions throughout every aspect of the research process. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Tatiana von Landesberger: Tatiana has extensive expertise in designing comparison studies and the visualization of meteorological data and has been instrumental in supporting throughout the research project, offering crucial feedback and constantly aiding in decision-making processes. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Human Strategies for Semantic-Based Colors Representing Multi-Label Classes

LAURA PELCHMANN¹, LENA JACOBI¹, SEBASTIAN BREMM² TATIANA VON
LANDESBERGER¹

¹ *University of Cologne*

² *Frankfurt University of Applied Science*

Abstract:

Visual analytics applications require the visualization of multi-label classes through colors, whether for data exploration during analysis or for examining the results of machine learning algorithms. While color schemes exist for single-label classes with semantic associations and for multi-label classes, it remains unclear how to create semantically resonant colors specifically for multi-label classes. To address this gap, we conducted two user studies and identified three effective strategies for creating color schemes for multi-label classes. We verified these strategies with meteorologists in a real-world use case. Additionally, we propose an algorithm for generating semantically resonant color schemes suitable for visualizing multi-label classes. We demonstrate the generalizability and practical utility of this approach through a further real-world data application.

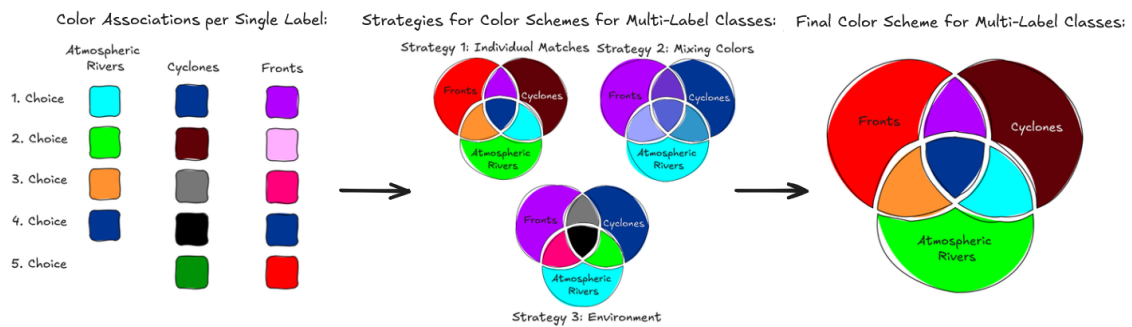


Figure 1: Example of semantic color assignments for a meteorological use case. **Left:** Expert color associations for single-label classes. **Middle:** Corresponding colors for multi-label classes using the three identified strategies from our user studies. **Right:** Final color scheme resulting from expert interviews.

1. Introduction

Visualizations and visual analytics systems are commonly employed to display class labels of classified data, such as the input or output of machine learning algorithms. In some cases, a class may be associated with multiple labels. These are known as multi-label classes (MLCs), multi-class, class mixtures, or combined classes. Examples of such scenarios include particles in the air that cannot be single-labeled due to close proximity or measurement uncertainties, such as rain and ice droplets or classes that exist simultaneously in the same location but different height, such as air and cloud motions that overlap on a map (Figure 1). Other examples include the classification of soil samples and possible mixtures of minerals, multiple land uses, or the presence of multiple animal species in one location. In all cases, the visualization should clearly communicate the meaning and co-occurrence of multiple labels.

Colors are often used in visualizations to show class labels [93, 136, 167, 174]. Choosing a suitable color scheme is challenging, as the colors of the classes must be unique and clearly distinguishable [36, 91, 98]. The focus on single-label classes and color schemes for hierarchies [104, 153] is not applicable to multi-label classes where all combinations may exist.

More recently, semantic color schemes have emerged that use inherent word-color associations (i.e., semantic) to enhance interpretability and represent the meaning of the data [2, 42, 67, 107, 135, 142, 174]. For example, Lexichrome [81] allows us to search for class labels and semantic color advice (e.g., cow – brown, rain – blue). The facilitated interpretability through association-based semantic color mapping may, in many cases, justify a slight deterioration in the pure distinguishability of the color values. This is offset by the increased interpretability, which ultimately improves the overall color scheme. However, semantic color schemes for multiple label classes are not yet available. It is not clear which color should be associated, for example, for "cow+pig" or "front+cyclone." Simple color mixing may not help to interpret the meaning of the data. This is especially needed in cases where the color transports the information about the data and several colors or glyphs are not possible.

A real-life example of color assignments, where semantic associations were considered to some extent for single-label classes but not for the corresponding multi-label classes, is shown in Figure 2a. The relationship between those classes and their combinations (displayed in Figure 2b) is not apparent to the scientists, thereby complicating

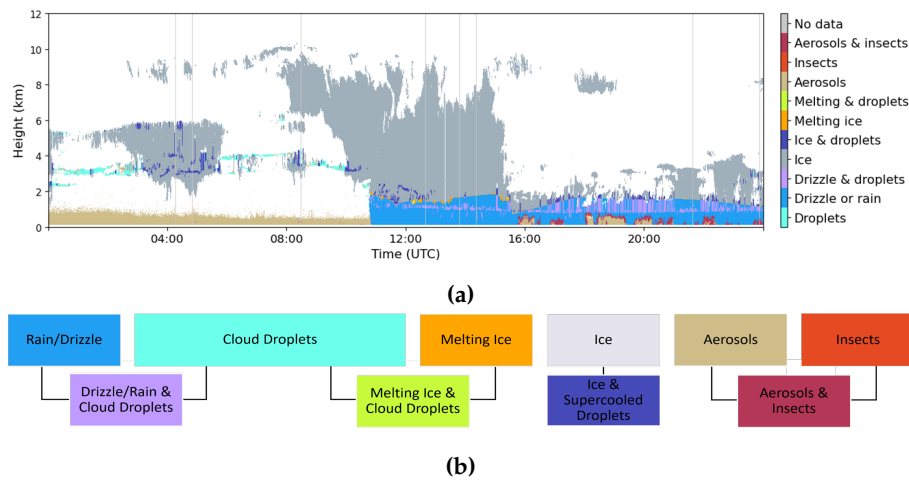


Figure 2: a) Depicts a meteorological time-height scatterplot that measures, classifies, and visualizes cloud particles up to a height of 12 kilometers according to the multi-label class color scheme over the course of 24 hours. b) Illustrates the used color scheme corresponding to the relationships of the displayed single- and multi-label classes.

potential analyses of the presented data. Thus, new strategies for identifying unique colors for MLCs are needed.

Our research question is:

What strategy should be used to create a multi-label class color scheme with strong semantic associations?

We conducted several user studies to derive human-centered strategies for identifying appropriate color combinations. In the studies, participants had to assign colors to different single-label classes and their combinations i.e., multi-label classes. The studies used different semantic datasets (animals, air particles, and weather conditions) to derive possible human-centered strategies across domains. These strategies were then applied to a meteorology use case for verification. Feedback from two domain experts confirmed the appropriateness of the strategies and their preference for color schemes in weather data. Overall, our research provides domain users with strategies for creating appropriate color schemes, along with detailed instructions to implement these strategies effectively.

2. Related Work

Colors for Single-label Classes Mapping categorical (i.e., single-label classes) or quantitative values to colors is a challenge in visualization research [26,95,114,127]. Suitable color schemes must ensure the usability of colors as data representations, their uniqueness, and optimal color assignment [22,36,91,98]. Distinguishable colors must be placed in a meaningful context to be well perceived by users, always depending on the number and types of data [54,57,106,154].

In addition, many categorical color schemes introduce the assignment of meaning to colors [142]. Semantic colors for categorical data (i.e., single-label classes) use inherent color associations to enhance interpretability [2,67,123,125,142,174] and represent information based on meaning rather than solely for aesthetic or decorative purposes [107]. More recently, language models have been used to help assign semantically resonant color

schemes [135]. However, these semantic colors have only been studied for single-label classes. Semantic colors for multi-label classes are lacking.

Colors for Compound Classes Related to MLCs is the color scheme for hierarchical (tree-structured) data presented by Tennekes and de Jonge [153]. The resulting qualitative color scheme assigns a unique color to each tree node, ensuring that there is no perceived order between nodes, that parent and child nodes are assigned similar colors, and that the depth of nodes is also represented by the color [153]. The colors can be evaluated according to Mertz and Kohlhammer [104]. This offers a wide range of options for applying colors to hierarchical data, potentially serving as a starting approach to identifying appropriate color scales for data with inherent semantic color associations and multi-label classes. The same is true for the dynamic assignment approach developed by Chen et al. [31] and weaving techniques presented by Hagh-Shenas et al. [56].

Another related topic is the visualization of intersecting sets [4]. Blake et al. [13] identified several ways to use color when creating Euler plots [79, 91, 105, 130, 162]. This is similar to [34, 78, 112]. They consider the choice of colors only as an aesthetic aspect to be considered when constructing Euler diagrams. Neither includes semantics and color distinguishability, especially for intersection regions.

3. User Studies

In order to answer our research question we formulated four hypothesis regarding the used strategies, perceived difficulties and factors influencing the color assignments. These hypotheses are composed of the previously mentioned results from studies on semantically associative color mappings [2, 67, 123, 125, 142, 174] and algorithms for coloring hierarchical graph structures [153]:

H1: In the case of MLCs, the average color (arithmetic mean) of the colors of the corresponding single-label classes is selected by most participants.

H2: A neutral color is selected for the MLCs comprising three labels.

H3: Colors for MLCs become more difficult to select as more labels are added.

H4: The first word in a combination of classes has a stronger impact on the color selection than the second word.

First, we conducted a pilot study in which participants could assign individual colors to predefined meteorological single-label and multi-label classes. After the pilot study, two studies were conducted for the experiment. One presented classes from biology in the form of nine different livestock classes (Cows, Pigs, Chicken, Sheep, Horses, Goats, Bees, Donkeys, and Ducks). The second study used nine different classes from meteorology, namely different particles that can be measured in the air (Rain, Drizzle, Cloud Droplets, Melting Ice, Ice, Aerosols, Insects, Supercooled Droplets and Dust). We chose datasets from different domains, all with semantic color associations, to ensure generalizability.

3.1 Pilot Study

In the pilot study, 81 participants were given the UW58 color palette, a set of 58 colors uniformly sampled in the CIELAB color space developed by Rathore et al. [125] used to find semantically resonant colors. Participants were also shown classes of a real meteorological use case [45, 72, 115]. They had to assign a color to each of the given classes (Droplets, Drizzle or Rain, Drizzle & Droplets, Ice, Ice & Droplets, Melting Ice, Melting

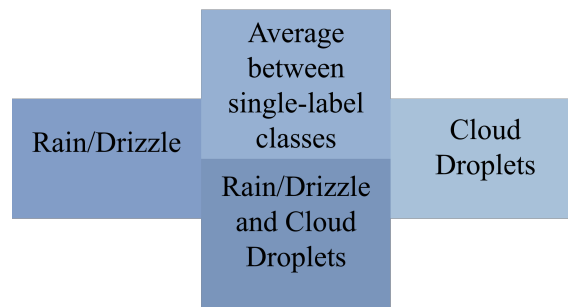


Figure 3: Shows the results for one class combination from our pilot study.

Ice & Droplets, Aerosols, Insects, Aerosols & Insects, No Data). A color picker allowed participants to assign an individual color.

The results indicated that, on average, darker colors were chosen for the multi-label classes (MLCs) compared to the average of the colors selected for the single-label classes (see Figure 3).

However, the selected colors were not suitable for a color scheme because the individually generated colors made the scheme partially indistinguishable. In the follow-up study, the color selection process was extended, but participants were limited to choosing from a set of predefined colors [115].

3.2 Main Studies

To improve the generalizability of our findings, we decided against using a fixed number of classes with predefined label combinations in our subsequent studies. Instead, we used a set of single-label classes, randomly selecting three for each participant. Participants had to assign a color from a predetermined palette to three of the respective classes individually, and then to each of the combinations of these classes representing the MLCs (example screenshots are shown in Figure 4). The procedure of both studies, using biological data and meteorological data, was identical.

Color Palette The procedure for creating the color palettes is based on the process used for the UW-58 colors [125]. To ensure that the colors are perceptually distinguishable, CIELAB coordinates were created in increments of 50 for each dimension. To ensure that this minimum distance between colors was maintained, a distance matrix was created showing the pairwise euclidean distances (ΔE) between all colors. In an iterative process, colors with a distance of less than 50 to at least one other color were then removed from the palette. This iterative process started with white (100,0,0), gray (50,0,0), and black (0,0,0). The final color palette for single-label classes consisted of 14 distinct colors, evenly distributed across the color space to ensure optimal distinguishability. To reduce the influence of the order of the displayed colors, they were displayed in the order shown in Figure 4 for a randomly selected subset of participants and in reverse order for the other subset. Subsequently, the average values of all pairwise combinations of these colors were calculated to assign colors for the MLCs, resulting in a color palette of 105 colors. The distance matrix identified all pairs of color whose distance was less than $\Delta E = 20$. In an iterative process, the colors that were too similar to the other colors, were removed from the palette and the distance matrix was updated accordingly. This process was continued until all remaining colors had a minimum distance of $\Delta E = 20$. Ultimately, 68 colors remained in the color palette available to study participants for selection for the MLCs. Again colors were displayed in the order shown in Figure 4 for a randomly selected subset of participants and in reverse order for the other subset.

*Please select the color you associate with the following term:

Cloud Droplets



You have selected the following color:

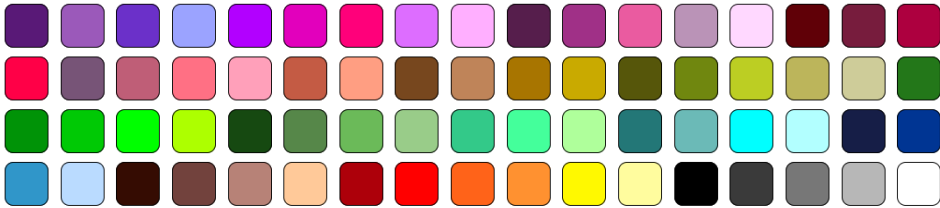
*How was it for you to assign a color to the term?

Choose one of the following answers

very easy easy neutral difficult very difficult

*Please select the color you associate with the following group of terms:

Drizzle&Cloud Droplets



You have selected the following color:

Figure 4: Shows screenshots of color assignment options for single-label (**top**) and multi-label classes (**bottom**) for the meteorological classes in our user study. Participants had to assign one of the colors provided.

Study Design The survey was conducted in LimeSurvey. Participants were welcomed with information on the study goal: *The aim of this survey is to analyze and compare different color associations. Main Task: In this study, you are asked to assign colors to different concepts.* To mitigate priming effects, we did not inform participants beforehand that the different color assignments would be combined at a later stage.

At the beginning participants had to disable the blue filter and night mode and increase the brightness to 100%, as described in [38] in order to allow the best color perception possible. A color blindness test was performed using the Ishihara tests [73]. Participants failing said color vision test were later excluded from the analysis.

In the main part, participants were presented with the names of three randomly assigned classes in sequence. They were asked to individually select the color they associated with each single-label class (L1, L2, L3) and to rate the difficulty of the task (see Figure 4). In a second step, participants were presented with their previously selected colors for each class and asked if they associated any additional colors with the single-label classes (L1.1, L1.2, ... L3.3). Up to three additional colors could be selected to verify during the analysis whether any additional associations influenced the color choice for the multi-label classes.

In the next part of the survey, participants were asked to select the associated colors for all possible MLCs combining two and then all three of the previously shown classes (see Figure 4). This happened without displaying the previously assigned colors for the single-label classes. Again, the difficulty of the task had to be rated. The color assignment ended with a control overview where participants were shown a summary of all their color assignments along with the corresponding classes. If colors were selected multiple times, participants were asked to adjust the scheme so that each color was selected only

Table 1: Shows pairwise independency tests for the color assignment of each label (**L**) as well as the alternative individual color assignments (**L1.1-L3.3**), **rows** to their respective MLCs (**columns**). Participants were initially allowed to assign a color multiple times (**top table**), and later had to assign unique colors while able to see the color scheme as a whole in a control overview (**bottom table**). Highlighted class combinations are significantly not independent (χ^2 -test, $p = 0.05$).

	MLC 1&2	MLC 1&3	MLC 2&3	MLC 1&2&3
L1	< .001	< .001	.017	< .001
L1.1	.004	< .001	.402	.298
L1.2	.381	.083	.065	.005
L1.3	.322	.718	.014	.161
L2	< .001	.133	< .001	.002
L2.1	< .001	.127	.030	< .001
L2.2	.493	.168	.127	< .001
L2.3	1.000	.988	.986	.999
L3	< .001	< .001	< .001	< .001
L3.1	.963	.666	< .001	.097
L3.2	.008	.935	< .001	.808
L3.3	.002	.214	.174	.042

	MLC 1&2	MLC 1&3	MLC 2&3	MLC 1&2&3
L1	< .001	.016	.337	< .001
L2	< .001	< .001	.012	.012
L3	< .001	.996	< .001	< .001

once. Participants were allowed to make changes even if there were no duplicate colors. Finally, participants were asked to describe their primary strategy for assigning colors. For this question, we provided multiple pre-formulated strategies (see Figure 5) and an option to formulate their own strategy. In the final questions, demographic information was collected, including gender, age, place of origin or previous residence, and any scientific agricultural or meteorological contact. An optional feedback box was available at the end of the survey.

Participants Participants were recruited from the university community. 113 individuals participated in the first study on biological data. 67% were female, 78% between the ages of 18 and 34, and 55% reported having a connection to agriculture. The average time to complete the survey was approximately 8 minutes. 56 individuals participated in the second study on meteorological data. 66% were female, 73% between 18 and 34 years old, and 32% reported a connection to meteorology. The average completion time was approximately 8.5 minutes. Participation in both studies was voluntary and participants were free to leave the experiment at any time.

4. Results

We first look at the statistical dependencies between the color assignments for the classes with a single-label and the color assignments for the classes with multiple labels. Then, we turn to the individual hypotheses and their evaluations.

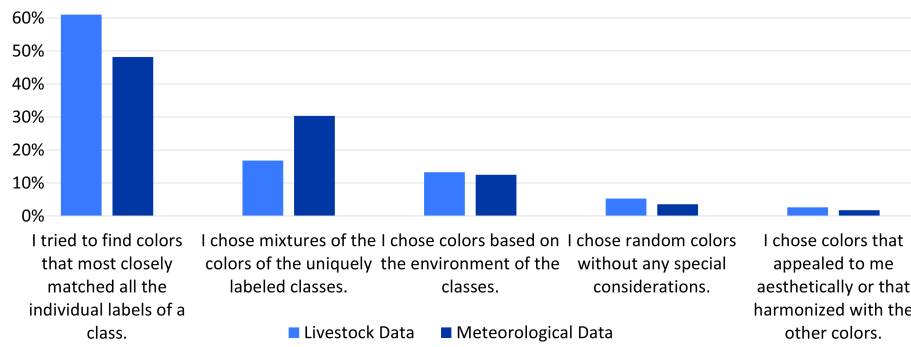


Figure 5: Shows the strategies participants used to assign colors to classes during our experiments and their relative frequencies.

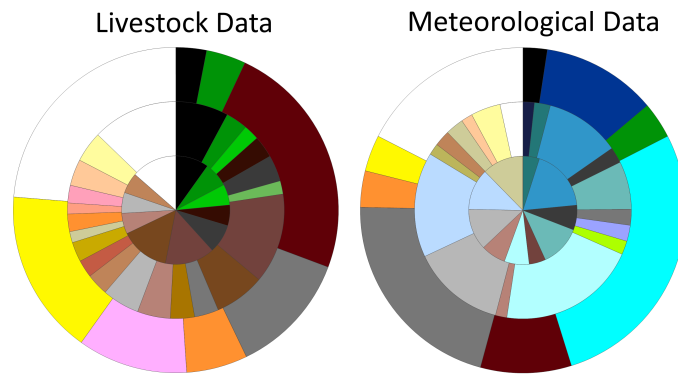


Figure 6: Shows the share of colors selected for the single-label classes (**outer layer**), classes with two labels (**middle layer**) and classes with three labels (**inner circle**) for both data sets during our user studies.

4.1 Statistical Evaluation

A total of 169 participants responded to the color assignments. We used a χ^2 -test to examine the independence between the selected colors for the single-label classes (L1, L2, L3) and the colors for the MLCs. Table 1 shows the results of these tests. Values that fall below the critical value of 0.05, rejecting the null hypothesis of independence are highlighted. The values clearly demonstrate that the colors chosen for the single-label classes, as well as their alternative associations, generally influence the selection of the MLCs. For the unique color selection in the control overview, the L2 label seems to have an influence on the color chosen for MLC 1&3. We suspect that this is due to the fact that both colors were shown in second position (L1, L2, L3 and MLC 1&2, MLC 1&3, MLC 2&3), which seemed to subconsciously imply a connection for participants.

4.2 Hypotheses

H1: In the case of MLCs, the average color (arithmetic mean) of the colors of the corresponding single-label classes is selected by most participants. The results show that only 21% of participants followed the strategy of using the average color values of the single-label classes to assign colors to the multi-label classes, leading us to reject this hypothesis. The most commonly used strategy, employed by 57% of all participants, involved selecting a common color for the multi-label classes within the single-label classes and assigning the highest possible priority to it. The strategies reported and their frequencies per dataset are shown in Figure 5. Only three participants

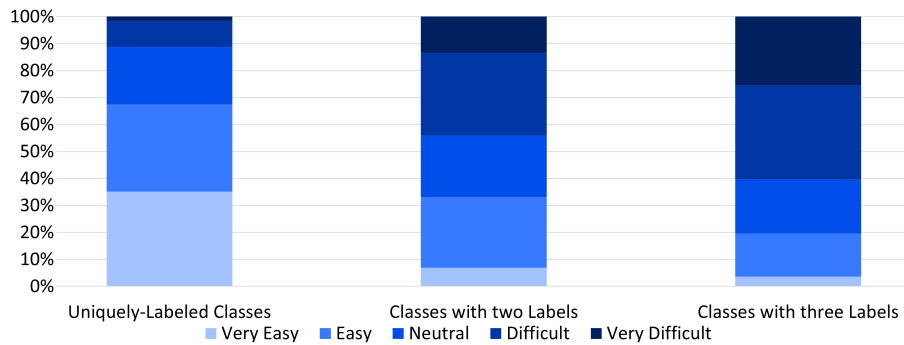


Figure 7: Shows participants' ratings of the perceived difficulty of the color assignments.

formulated their own strategies. Of these, two reported attempting to use a combination of the provided strategies. One participant indicated that their choice of colors was influenced by previous research.

This result suggests that for the creation of color schemes for multi-label class visualizations, we should focus on collecting multiple possible color associations per single-label class and then assign the highest possible priority to every class.

H2: A neutral color is selected for the three-label classes.

Participants' color assignments for classes with different numbers of labels per study are shown in Figure 6. In the innermost circle, not only neutral colors were chosen. So we have to discard our **H2**. However, the range of colors chosen decreases from the single-label classes in the outer ring of the circle to the multi-label classes (MLCs) in the inner circle, thereby reinforcing agreement among participants. It seems that the choice of colors is based on the general association of the topic of the data, i.e., brown and green for farm life and blue and gray for meteorology.

This result suggests that for the color selection of multi-label classes, a color assignment might be sufficient that is generally close to the color association of the data topic.

H3: Assigning colors to MLCs becomes more difficult as more labels are added.

While 35.1% of the color assignments for each single-label class were rated as very easy, only 6.9% of the assignments for the classes with two labels were rated as very easy, and only 3.6% of the assignments for the classes with three labels. Conversely, 1.6% of the color assignments for each single-label class were rated as very difficult. This was 13.4% for the classes with two labels and 25.4% for the classes with three labels.

We can therefore confirm **H3** with our observations. The difficulty of color assignments increases with each label added to a class.

Since assigning colors to the classes with three labels is the most challenging, we recommend beginning with this color assignment. Subsequently, proceed to assign colors to classes with two labels, followed by the single-label classes in the subsequent steps. This approach is advised because participants found the assignment for single-label classes to be the easiest, and a broader collection of color associations for these single-label classes exists.

H4: The first word in a combination of classes has the strongest impact on the color selection.

We measured the distances between the colors chosen for the single-label classes and the colors chosen for the classes with multiple labels. These distances were categorized as follows:

- Neither of the two colors is closer to the assigned color for the MLCs, it is rather represented by the average color value.

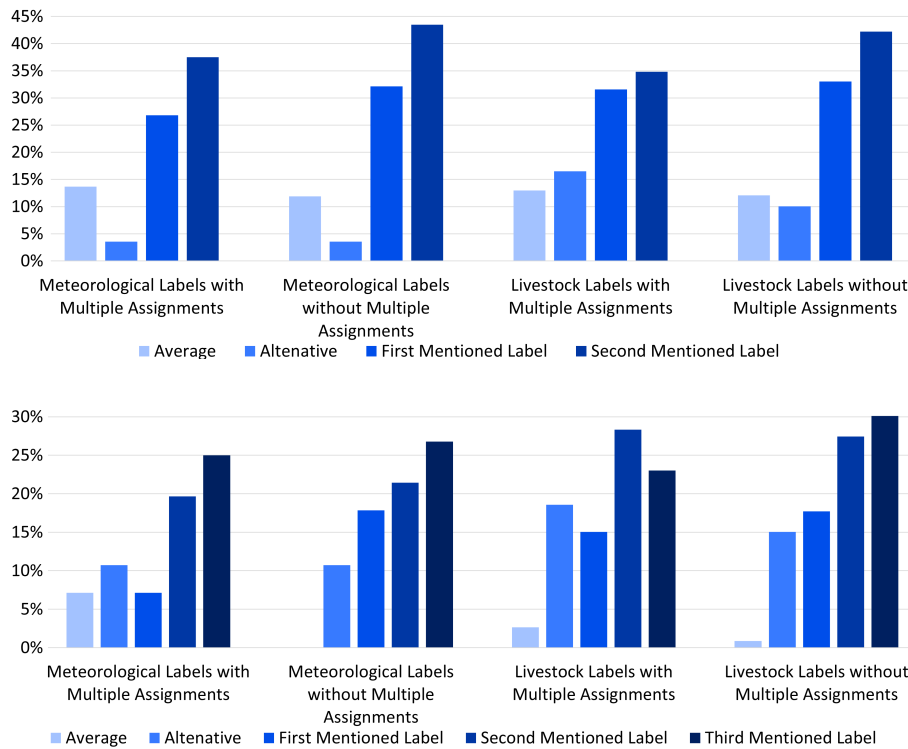


Figure 8: Shows the similarities between the colors selected for the single-label classes and the color selected for the corresponding classes with two labels (**top**) and three labels (**bottom**). In seven out of eight categories the highest similarity is between the color chosen for the MLCs and the last mentioned single-label. **Multiple Assignments** allowed participants to assign the same color more than once, corresponding to the first part of the study. This differs from the unique color assignments from the control overview (**without Multiple Assignments**).

- One of the colors marked as alternative associations is closest to the color for the class with multiple labels.
- One of the colors assigned to the single-label classes is closest to the color chosen for the MLCs.

Figure 8 shows that we need to discard our **H4**. In seven out of eight categories we examined, the color assigned to the last mentioned single-label class was closest to the color chosen for the multi-label class, not the first one.

This result implies that if there is no color for a multi-label class that occurs as an association in all single-label classes, the greatest focus while assigning a fitting color should be on the last-mentioned label.

5. Expert Evaluation Use Case

To test the usefulness of the defined strategies, we adapted a sample color scheme in meteorological visualization (Figure 9). It allows the examination of precipitation originating from atmospheric rivers (AR), cyclones (C), or fronts (F), or their combinations. According to domain experts who sought our advice, the original color scheme needed improvement due to poor differentiability. The new colors were expected to be easily distinguishable while remaining both intuitive and relevant to the different classes.

Color Associations per Single-Label In expert interviews with two meteorologists, we asked for the color associations of the three single-label classes (AR, F and C). Up to four

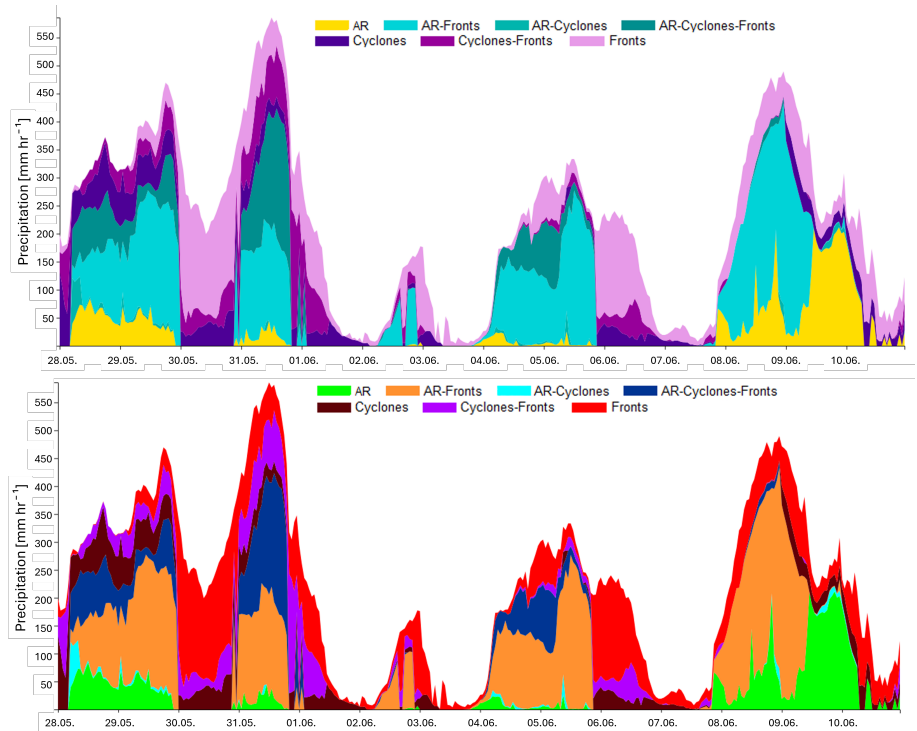


Figure 9: Top: Shows the originally used visualization by meteorological scientists. The data displayed are used to measure precipitation in connection to atmospheric rivers, fronts and cyclones. The color scheme should be adjusted according to the wishes of meteorological experts due to issues with the color distinguishability. **Bottom:** Shows our color scheme suggestion resulting from Strategy 1, following the expert interviews. This color scheme primarily concentrates on the hues chosen based on color associations. Once these hues are established, further refinements can be made to other aspects, such as saturation.

additional colors could be assigned. There was a clear consensus on associations of bright signal colors for atmospheric rivers, associations of cold and warm, or a mixture thereof, for the fronts, and of somewhat darker colors for the cyclones (Figure 1).

Color Schemes Following the Human Strategies We developed color schemes based on the observed strategies of the previous two surveys on the basis of selected single-label color associations. Each strategy lead to a different result, which were then discussed with the domain experts. The three schemes rated best by the experts are in Figure 1.



Strategy 1: Colors closely matching individual classes

A balance is achieved where the colors for the multi-label classes represent as much of associated colors for each included single-label classes as possible. In this case the alternative associations for atmospheric rivers were given a preferential role for the two-label classes. This color scheme is appropriate for dichromatic viewing (protanopia, deuteranopia, tritanopia) and has distinguishable colors.



Strategy 2: Mixtures of colors

The strongest associations were chosen for the single-label classes, and average color values were used for the MLCs. Although both experts found it sufficient for the purpose, the color scheme is not appropriate for dichromatic viewing (inappropriate for protanopia & deuteranopia, appropriate for tritanopia) and the colors are generally not

well distinguishable. This approach would also result in a similar color scheme as the colors originally used by the meteorologist ([Figure 9 Top](#)). An attempt to make it more distinguishable by using black for the 3-label class was rejected by the experts because of the lack of association of black with the three-label class.



Strategy 3: Colors based on the environment of classes

The strongest associations were selected as a base, and for the two- and three-label classes, associations were attempted for a real-world environment. For our use case, this was quite difficult because the environment – i.e., sky, rain, and clouds – is similar for all combinations. The color scheme is appropriate for most dichromatic vision (appropriate for protanopia & deuteranopia, inappropriate for tritanopia). The colors are distinguishable. The experts were not completely convinced by the color black, which represents the class with three labels. However, this color scheme is useful.



Final Color Scheme

Both experts named two favorites for the final color scheme: Expert 1 preferred the result of strategy 1 slightly more than the result of strategy 2. Expert 2 also preferred the same results, but prioritized the result of strategy 2 above the result of strategy 1. Given that the outcome of strategy 2 was less favorable in terms of color distinctiveness and color vision deficits, and also resembled the scheme intended for modification, we ultimately opted to utilize the color scheme resulting from strategy 1. [Figure 9](#) shows a version of the visualization using said color scheme example.

For this resulting color scheme, it is important to note that established color concepts from visualization and perception research might concede a poor distinguishability of the colors [3, 74, 168]. However, our interviews reveal that the semantic color association of scientists might hold greater importance, both for single-label classes and a semantically based logic for multi-label classes, compared to a purely perception-based distinguishable color scheme. This may hold different levels of priority for each scientist and every use case, which can be assigned individual importance in every resulting color schemes.

6. Algorithm for the Design of a Multi-Label Class Color Scheme

Based on the results of the quantitative study and qualitative expert evaluation, we have developed instructions for designing an appropriate custom color scheme for multi-label classes. These instructions follow Strategy 1 from the example mentioned above but combine it with Strategy 2 and can be found in [Algorithm 3](#).

Algorithm 3 Algorithm for Assigning Colors to Multi-Label Classes

Input:

A set of single-label classes: $C = \{C_1, C_2, \dots, C_n\}$ and a prioritized list of selected colors $S_i = [c_{i1}, c_{i2}, \dots, c_{im}]$, where c_{i1} is the highest priority color and c_{im} is the lowest priority color

Step 1: Color Collection

for each single-label class C_i **do**

 Prioritize at least three color associations to create the prioritized list S_i

end for

Step 2: Assign Color for the Three-Label Class

for each combination of three single-label classes (C_i, C_j, C_k) **do**

Check for Common Color:

 Compute the intersection of the sets of selected colors: $S_{\text{common}} = S_i \cap S_j \cap S_k$

if $S_{\text{common}} \neq \emptyset$ **then**

 Select the highest priority color from S_{common}

 Assign this color to the three-label class

 Exit

else

Assign Default Color:

 Compute the usage count of black and gray in the single-label classes

if the usage count of black or gray is zero for the single-label classes **then**

 Assign black or gray to the three-label class

else

 Assign a color not highly prioritized for any of the single-label classes

end if

 Exit

end if

end for

Step 3: Assign Colors for Two-Label Classes

for each combination of two single-label classes (C_i, C_j) **do**

Check for Common Color:

 Compute the intersection of the sets of selected colors: $S_{\text{common}} = (S_i \cap S_j) \setminus S_k$

if $S_{\text{common}} \neq \emptyset$ **then**

 Select the highest priority color from S_{common}

 Assign this color to the two-label class

 Exit

end if

Create Color Mixtures:

 Identify high-priority colors that are most similar to each other

 Use a color mixture from these high-priority colors

 Exit

end for

Step 4: Assign Colors for Single-Label Classes

for each single-label class C_i **do**

 Identify the set of multi-label classes that involve C_i

 Assign the highest priority color from S_i that is not used in any of the multi-label classes involving C_i

end for

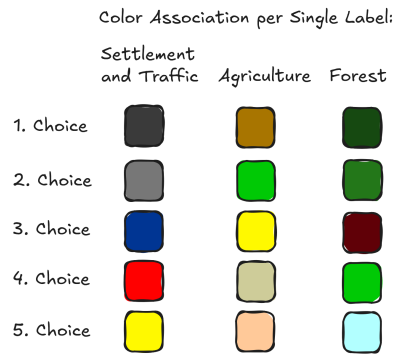


Figure 10: Shows the color associations of an interview partner for the single-label classes of land use in Germany.

7. Application

We now aim to apply the algorithm to a practical application. For this purpose, we utilized a simplified dataset from the Federal Statistical Office of Germany on land use across German federal states [1]. Our objective is to visualize the dominant land use class(es) for each federal state on a map.

To achieve this, we conducted an interview with an independent participant to elicit color associations for the three predefined classes: "Settlement and Traffic," "Agriculture," and "Forest." Subsequently, we follow the algorithm's procedural steps to derive an appropriate semantically resonant color scheme.

Step 1: Color Collection

Step 2: Assign Color for the Three-Label Class

Since there is no common color for all three single-label classes, we count the assignments of black and gray.

Since gray was chosen for the color associations, but black was not, we assign black to the three-label class.

Step 3: Assign Colors for Two-Label Classes

Settlement and Traffic and *Agriculture* have a common color not being chosen for Forest.

Hence, the yellow is chosen for this two-label class: 

Settlement and Traffic and *Forest* have no common color. Hence, we mix the dark gray and the dark brown for the two-label class: 

Agriculture and *Forest* have a common color not being chosen for Settlement and Traffic. Hence, the light green is chosen for this two-label class: 

Step 4: Assign Colors for Single-Label Classes

Settlement and Traffic has its first color choice used for the color mixture for the two-label class with forest. Hence, we use the second choice: 

The first choice for *Agriculture* has not been used for any of the multi-label classes. Hence, we use that first choice: 

The first choice for *Forest* has not been used for any of the multi-label classes. Hence, we use that first choice: 

Step 5: Final Visualization

This results in the final color scheme used for the map visualization.

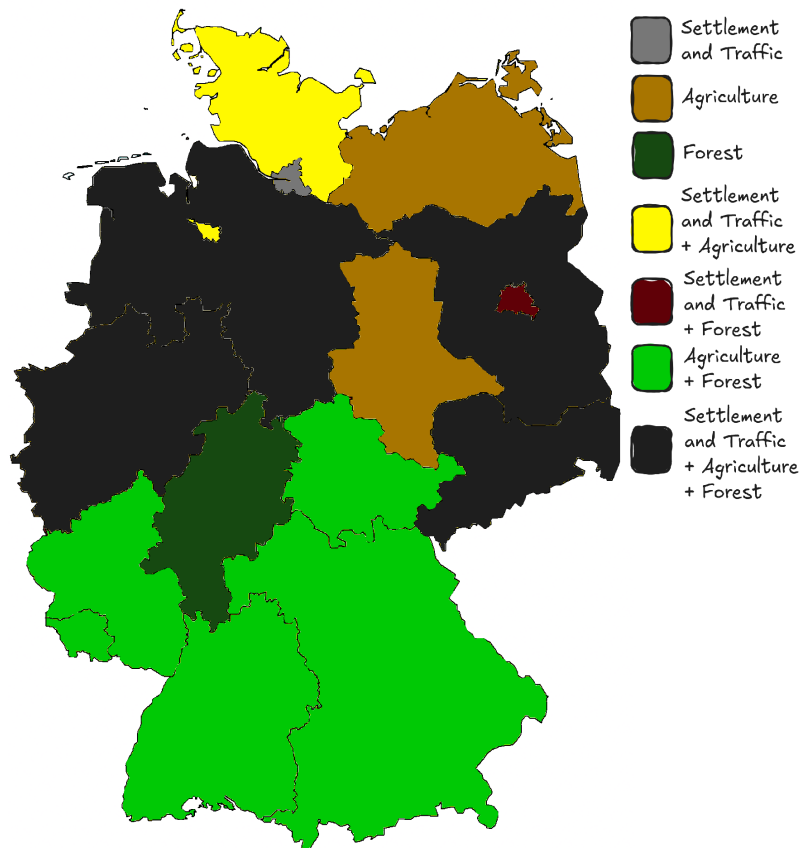


Figure 11: Shows the primary land use in Germany per federal state.

8. Discussion and Future Work

We investigated human strategies for creating semantically resonant colors for multi-label class data and developed a set of instructions for constructing an individually suitable color scheme. Our research indicates that most individuals try to implement both the semantic associations to class labels and the connections between classes with identical labels when generating such color schemes. However, the semantic color association appears to have a more substantial influence. Our expert use case illustrates the process of creating such a semantically resonant color scheme. It shows the pros and cons of different strategies. The expert interviews show that it is worthwhile to create color schemes according to different strategies of semantic associations of multi-label classes. Our subsequent application of the algorithm to an independent dataset demonstrates the implementation and practicality of said algorithm.

It is worth noting that while the process was designed and tested with 3-label classes, the algorithm itself could also be applied to 4-label classes. The steps would consequently need to be expanded. However, it is important to emphasize that such a number of single-label classes and their corresponding combinations would significantly exceed the limit of seven distinguishable colors for a categorical color scheme mentioned in the related work. It is important to note that the study design, which first queried color associations for single-label classes and subsequently for corresponding multi-label classes, may have introduced potential priming effects for the color choices for the MLCs. Our work is

primarily focused on identifying various strategies for creating color schemes. In future studies, an evaluation of the resulting visualizations would be desirable, as well as exploring possibilities for adjusting saturation, brightness, and transparency. For future research, it might also be interesting to unify and automate the different strategies and build a tool to automatically create semantic color schemes for the scientific community. It is also interesting to investigate whether individuals tend to choose darker hues than the mathematical average. This is a question that should be addressed in future studies.

9. Acknowledgments

We would like to extend our sincere thanks to all the participants who contributed to this study. We also thank Felix Müller for his invaluable help in presenting the meteorological data for this research paper. This work was partially supported by the BMBF WarmWorld Project (01LK2204A), the DFG SANE Project (527250730) and the BMBF Risk Principe Project (01ZZ2323J).

3.2 Discriminability of Node Colors in Node-Link Diagrams: Graph Comparison

3.2.1 Introduction

The following paper investigates factors influencing the detection of color differences across graphs. In a study, participants were shown two identical graphs above each other, in which only individual nodes or edges were colored differently. They had to identify these color differences within a time limit.

3.2.2 Factors Influencing Visual Comparison of Colored Directed Acyclic Graphs

This paper [55] was published and presented at the Vision, Modeling, and Visualization Conference (VMV), 2023.

Graniczkowska, C., **Pelchmann, L.**, von Landesberger, T. and Pohl, M. (2023). Factors Influencing Visual Comparison of Colored Directed Acyclic Graphs. *Vision, Modeling, and Visualization*. The Eurographics Association. doi: [10.2312/vmv.20231232](https://doi.org/10.2312/vmv.20231232).

I am the secondary author of this publication. In this role, I was responsible for the analysis, as well as the writing and publication of the work. The individual contributions of myself and my co-authors to this publication are outlined below according to the Contributor Roles Taxonomy (CRediT):

Cynthia Graniczowska: Cynthia conducted her master's thesis on factors influencing the visual comparison of directed acyclic graphs, guided by Tatiana and me and supervised by Margit. In this regard, she has notably contributed to the concept, conduct and analysis of the user study. Moreover, she proofread the final paper submission. Contributor Roles: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing.

Laura Pelchmann: I contributed to the study's conception process and helped determine the colors for the graph nodes. Following that, I performed an extended analysis of the collected study data and wrote the research paper. Contributor Roles: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft.

Tatiana von Landesberger: Tatiana has extensive expertise in designing graph comparison studies and has been contributed in guiding Cynthia and me throughout the research project, offering crucial feedback and aiding in decision-making and writing processes. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing.

Margit Pohl: Margit has given significant insights into psychological studies and their design, pertinent previous literature, and the appropriate variation of stimuli. Additionally, she consistently provided feedback and addressed questions about all aspects of the research process. Contributor Roles: Conceptualization, Methodology, Supervision, Writing – review & editing

Factors Influencing Visual Comparison of Colored Directed Acyclic Graphs

CYNTHIA GRANICZKOWSKA¹, LAURA PELCHMANN², TATIANA VON LANDESBERGER²,
MARGIT POHL¹

¹*TU Wien*

²*University of Cologne*

Abstract:

This paper presents a comprehensive investigation of the factors that influence visual comparison in colored node-link diagrams. We conducted a user study in which participants were asked to identify differences in pairs of directed acyclic graphs (DAGs) under time constraints. Previous studies focused on the perception of differences in node-link diagrams without coloring. Our results show that the individual coloring of nodes and edges significantly affects the detection of differences. We were able to confirm previous results, such as the influence of graph density, and also found that uniform coloring in certain areas of the graphs plays an important role in finding differences. Consequently, the results of this study hold potential for developing better comparative visualizations for diverse applications, such as finance or biology.

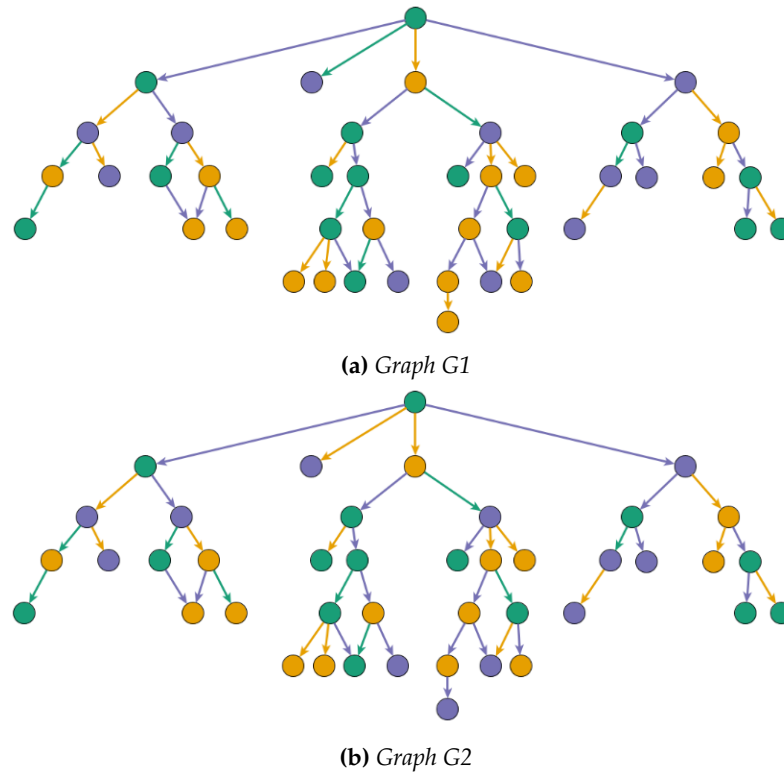


Figure 1: A pair of graphs shown to our participants including minor differences in the coloring of edges and nodes shown in Graph G2, that were to be spotted and marked by the participants.

1. Introduction

Visually comparing directed acyclic graphs (DAGs) represented as node-link diagrams is common in many disciplines, such as finance and biology e.g. [39, 64, 87, 163]. The comparison of these graphs relies heavily on the finding of differences, e.g., finding changes in nodes and edges, such as added, or deleted nodes and edges, or changes in the color of nodes and edges. This analysis often happens under time pressure.

The aspect of color can also be used in such DAGs. For example, different categorical attributes can be represented with color-coded nodes. The application areas range from the categorization of different actors in the financial sector to the position classification of nodes within a gene ontology database, e.g. [46, 84] and require the perception of differences and similarities in the underlying data features [161]. Therefore, the question of the influence of different visual variables on the detection of these differences and similarities is an important aspect of related research on human graph comprehension [50, 163, 166]. To this end, many studies have been conducted in recent years, and various insights have been gained into the influence of edge intersections, shape, density, and other aspects of DAGs [5, 165, 166].

Since color classifications did not play a role in previous studies, but could influence the comparison of DAGs, we decided to investigate the influence of colored nodes and edges on the comparison of different DAGs in a new user study (see Figure 1). In order to focus the study more closely on the coloring, the structure of the graphs that are compared with each other does not change. Changes only appear in the coloring of individual nodes and edges. In order to still be able to compare the effects of coloration with

previous results, the factors we examined are closely aligned with those of Ballweg et al. and Wallner et al. [5, 165, 166]. Our study answers the following research questions:

- Do the variable color and the factors white space, position, uniform/multicolored background, graph size and edge length influence the recognition of differences in DAGs under time constraint?
- Do the variable color and the factors white space, position, uniform/multicolored background, graph size and edge length affect the sequence of perceived differences of DAGs? Are some variables perceived earlier than others?

Our results show that the influences of density/white space, already found by Wallner et al. also have a significant impact on the perception of differences in colored DAGs under time constraints. Furthermore, our results indicate that a uniformly colored environment of nodes and edges, or the lack thereof, has an important influence on the detection of differences. In addition, we found indicators that differences concerning uniformly colored surroundings and long edges were more likely to be found first than others. Hence, it might be advisable to highlight differences in multicolored environments and in connection to short edges to make it easier for users to identify them.

The next section provides a summary of related work. This is followed by a description of the study design in Section 3, and a description of the quantitative results obtained in Section 4. These are further discussed in Section 5. Finally, an outlook on future work building on the results described here is given.

2. Related Work

We shortly review related work concentrating on the areas of influencing factors of graph comparison, color, as well as further work on graph perception in general.

2.1 Influencing Factors and Strategies of Graph Comparison

Our study is closely related to and builds on the research of Ballweg et al. [5] and Wallner et al. [165, 166].

Ballweg et al. [5] conducted a sorting study with 20 participants in search of factors influencing the perception of similarity of DAGs. In this study, 69 different graphs with 6-9 nodes each had to be sorted according to similarity and reasons had to be given. Both a quantitative and a qualitative evaluation showed that the most important comparison factors for the participants were the number of levels, the number of nodes in each level and the outer shape of the graphs. Edge crossings, on the other hand, did not seem to have a significant effect on participants' perceptions. However while this study concentrated on similarity, we focused on the perception of differences.

Subsequently, Wallner et al. published [166], a first qualitative study on the perception of differences in directed acyclic graphs. In contrast to Ballweg et al., this study focused on differences rather than similarities. According to the authors, this leads to different results. When asked for similarity, users tend to concentrate on global features (e.g., outer shape), but when they are asked to identify differences they focus on more local features (e.g., edge crossing). In addition, the authors argue that possibly different layouts (orthogonal vs. layered) and nature of tasks may afford different strategies of analysis.

The authors found evidence that fewer edge crossings make graphs easier to read, while crossings that do not exist in the original and are created during manipulation seem to be visually dominant. The results also suggest that greater symmetry in the

layout increases the comparability of the graphs. Finally, the shape of the graphs also seemed to be a relevant aspect. The study also found that the density of the graphs and the amount of white space within the graph are closely related, and suggests that changes should be made in regions that are less dense or even empty.

In a follow-up study, Wallner et al. looked more closely at the factors that seemed to be most important in their qualitative study: shape, density, and edge crossings [165]. This time, participants were given a limited amount of time to detect differences in the same graph pairs. The authors argue that the effects of time constraints have not been investigated extensively so far. There is some empirical evidence that time constraints cause study participants to process information in a more shallow manner and perceive mainly the more important features [151]. Working under time constraints also increases the cognitive workload of participants [14,63]. The authors assume that the introduction of time constraints changes the cognitive strategies of the users.

After each pair of graphs, participants were asked to rate the certainty and difficulty of finding differences in that pair of graphs. The shape of the graph helped in most cases to detect this change before other changes. However, this aspect did not seem to be critical in determining whether a change was detected or not. It is these results found by Wallner et al. in [165,166] that we now want to investigate in DAGs whose nodes and edges are colored, since coloring was not considered as a factor to be investigated in the previous studies. Therefore, the factors we investigate are very similar to those discussed above. A correlation was found between certainty and difficulty, between differences found and certainty, and between differences found and difficulty. Although changing the outer shape of the graph did not have a statistically significant effect on the frequency of differences found ($\chi^2(1) = 2.48, p = 0.115$), Wallner et al. found that participants found these differences faster.

A summary of the results listed here can be found in Table 1

Type of Dif.	[5]	[166]	[165]
Shape	yes	yes	no
Edge Crossings	no	yes	yes
Density	no	yes	yes

Table 1: Summary of what types of differences were found to have an impact on the comparison of participants.

2.2 Color

The usage of color in visualizations has been studied quite extensively. We may distinguish between two types of investigations, on the one hand investigations addressing how to adopt colormaps in visualizations to represent numerical values faithfully [95] [26] [127]. On the other hand, there are investigations on the usage of color to distinguish between different categories of objects. This area of research is more relevant for the investigation described here.

Szafir ([152]) argued that most of the traditional investigations about color perception are based on laboratory studies studying differences between larger areas of color under perfect viewing conditions. She investigated the usage of colors under more realistic conditions and in the context of visualizations with fairly small areas of color (scatterplots, line charts, bar charts). In such cases, it is much more difficult to distinguish between elements with different colors. Smart and Szafir ([148]) also point out that the perception of various features of visualization elements (shape, size, color) influence each other. One

result of their research that is especially relevant for using color is that shape affects the perception of color differences.

Jianu et al. [75] investigated the influence of larger areas of color on the perception of groups of elements. Saket et al. [133] were able to show that the usage of large areas of color also increases enjoyment of use of visualizations. In our current research, we do not use larger areas of color to distinguish between different clusters of nodes, but this is certainly a possible area for future research.

2.3 Graph Perception

Burch et al. published an important state of the art work on the empirical user evaluation of graph visualizations, in which they identify the white spots in this field and sketch ideas for future research directions [27]. To study the perception of visual properties of graphs, many works focus on single graphs. Lin et al. [90] investigated different layout methods and determined which nodes of a graph are visually more important based on the degree of the nodes, the number of neighboring nodes, and the number of edges and edge crossings in a surrounding area.

Mariott et al. [101] designed a study in which participants had to draw graphs from memory. They concluded that symmetry, colinearity, and orthogonality are important properties when recalling a graph. Yoghoudjian et al. [173] surveyed the literature on human-centred experiments to understand how, in practice, different features and characteristics of node-link diagrams affect visual complexity, while Soni et al. [147] conducted an experiment to investigate whether the perception of graph density and the average local clustering coefficient can be modeled using Weber's law and Quadri and Rosen investigated the influence of visual density on cluster perception in scatterplots [122]. The effects of curves and edge crossings on graph perception have also been addressed, by Huang et al. [68,69]. The question of how to support comparison processes of visualized data is a very active area of research. Gleicher et al. [50] believe that the development of future comparison tools will be facilitated if the understanding of comparison in general is developed. Although there are many different systems and approaches, they found that the basic types of visual comparison techniques include: juxtaposition (showing different objects separately), superposition (overlying objects in the same space), and explicit encoding.

In recent years, several systems have been developed to support comparison processes of visualized data. Chen et al. [32] developed a system to compare multi-item data streams. Joos et al. [76] developed a system to support comparison processes in VR. Pister et al. [117] developed a system for querying social networks and comparing the resulting selections. Researchers have also addressed more general questions. L'Yi et al. [97] studied the influence of different layouts on comparison processes. They compare juxtaposition, superposition and explicit encoding and argue that all these methods have strengths and weaknesses. They also provide recommendations when to use the approaches and how to counteract negative effects. Gaba et al. [47] describe how natural language has been used to describe comparisons and the implications for visualization. They found out that bar charts were preferred by many users as the most appropriate kind of visualization for comparison processes. In addition, they derived guidelines for supporting comparison processes, e.g., that these processes should be supported by interaction and text. Xiong et al. [171] also investigated natural language expressions and what implications these expressions could have for the design of visualizations to support comparison processes. Based on their research, they argue that users prefer to conduct comparison processes with visually aligned bar charts. They also developed guidelines for the design of bar charts to support comparison processes. Nevertheless, they point out that other

types of visualizations should be investigated. Pandey et al. [113] conducted an experiment to investigate the perception of similarity of scatterplots. Participants had to group plots according to their individual understanding of similarity. They identified key concepts for the perception of similarity.

The work of Bridgeman et al. [25] on similarity measures for graph drawing suggests that both absolute and relative point positions are indeed important for perceiving similarity.

3. Study Design

Based on the results of Ballweg et al. and Wallner et al. [5, 165, 166] as well as other studies listed in the related work, we decided to re-examine the aspects of shape and density/white space of a graph under the aspect of coloring. We included the major variables investigated in previous studies, but omitted edge crossing as an influencing factor to keep the number of variables fairly small and manageable. Additionally, based on the factors investigated and not tested so far in the graph perception research, we chose to also investigate new factors. Therefore, in this study we also considered graph size, different colors in uniformly colored environments, and edge length (the Stimuli-Factors we considered are further explained in Section 3.4), and came up with the following six hypotheses:

- H1 Differences are more likely to be found in small graphs.
- H2 Differences are more likely to be found on the outside.
- H3 Differences are more likely to be found where white space is.
- H4 Differences are more likely to be found on the background of uniformly colored edges and nodes.
- H5 Differences are more likely to be found in changes to longer edges.
- H6 Differences with certain characteristics are found significantly earlier than others.

The survey was set up in "Limesurvey" to investigate how the perception of graphs is changed by the coloring of nodes and edges, in order to find information on how potential class information in graphs can be optimally represented by coloring. Different pairs of directed acyclic graphs were displayed to the participants in the survey, which they then had to compare. The participants were required to mark the perceived differences between the graph pairs using a drag and drop functionality. They placed markers on the selected locations. For more details on the procedure, please refer to Section 3.2. Figure 3 provides an example of the study setup.

We decided to only use unlabeled node-link diagrams to avoid the influence of intervening variables (e.g., label semantics) and to be able to identify the relationships between well-defined variables (position and color). Using labeled node-link diagrams would make it difficult to decide whether the results are due to the use of color or to the specific character of labels.

The following subsections describe in more detail the aspects of coloring, the exact procedure of the study, as well as the study participants and the factors to be examined in the study.

3.1 Color Coding

Since individual coloring of nodes and edges is a new factor in this type of graph perception research, the number of colors to be used was set to three. Three colors offered

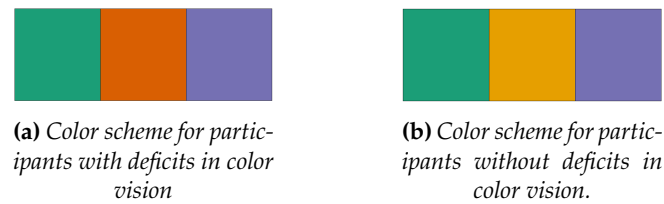


Figure 2: Two different color schemes used in the Study.

us enough variability in color assignment for the study without creating too much complexity. With the help of ColorBrewer, the colors to be seen in Figure 2a were chosen to ensure safe color vision for people with color vision deficits. However, after testing with participants without color vision deficits, we decided to use a slightly modified scheme for them, as they found the tested colors less distinguishable. For them, the scheme in Figure 2b was used.

3.2 Procedure

The study consisted of 16 graph pairs. For comparability with previous studies, we used a very similar data set as Wallner et al. [165]. For each question, a base graph and the altered version were displayed at the same time, one on top, the other one below (an example is shown in Figure 1). The differences in the bottom graph influenced the colors of the nodes and/or links, but did not change the structure of the graph at all (see Section 3.4 for more details). Participants had to mark the perceived differences by dragging and dropping a marker to the desired location, nodes and/or links, in the bottom graph (as shown in Figure 3).

The graph pairs were presented in a semi-random order. We had 6 different sequences to avoid showing similar changes one after the other. We started with two test questions to familiarize the participants with the system (testing the drag and drop functionality and adjusting the zoom factor of the screen). The first test question had no time limit, while the second had a time limit of 3 minutes.

For the quantitative study, a time limit was set for each pair of graphs in order to more quickly filter out the differences that caught the participants' attention and to avoid too precise a systematic search. This was 40 seconds for the small graphs and 60 seconds for the large graphs. The time limits were determined in a small pilot study.

In addition, after each pair of graphs, participants were asked how confident they were that they could find all the differences and how difficult it was to find the differences. Possible responses were measured on a 5-point scale ranging from 1: very certain/easy, to 5: very uncertain/difficult.

3.3 Participants

For this visual survey, we selected computer science students as participants.

In total, we received 41 complete responses from 11 females and 30 males, with four of the participants having deficits in their color vision, hence working with the alternative color scheme displayed in Figure 2a. The average age of the participants was 25, with the youngest being 22 and the oldest being 29. On average, participants rated their familiarity with visualizations as 2.5 on a 5-point scale (1: very familiar, 5: very unfamiliar). The average time to complete the survey was 22 minutes.

3.4 Stimuli-Factors considered

For each difference in the coloring of individual edges and nodes of a graph pair we defined five factors, that were systematically varied among these graphs and are further explained in the following paragraphs:

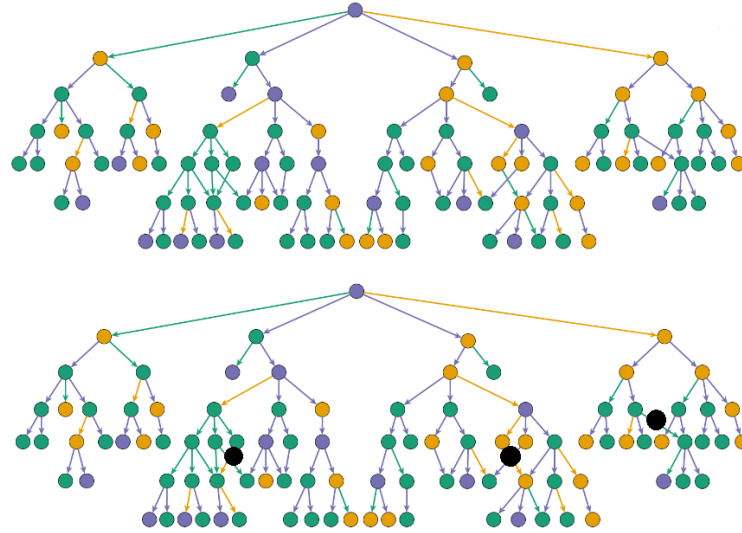


Figure 3: Two exemplar large graphs shown to our participants. Differently colored nodes and links had to be marked in the bottom graph.

- graph size (small/large) (H1)
- position (not impacting the shape (inside)/impacting the shape (outside)) (H2)
- white space (yes/no) (H3)
- color (uniform/multicolored) (H4)
- edge length (no edge/short/long) (H5)

The factor *graph size* could be either "small" (see Figure 1) or "large" (see Figure 3). Eight graphs were considered "small" and had between 40-56 nodes and 44-61 edges, while the other eight graphs were declared "large" and had between 97-104 nodes and 104-124 edges.

Looking at the factor *position*, to define whether a difference is on the inside or outside of a graph, we specified that any difference that does not affect the shape of the graph is considered to be on the inside. This factor is strongly related to the shape factor underlined by Wallner et al. Unlike the previous studies, however, nodes and their associated edges are not added or omitted, but merely recolored.

As for the *white space* factor, we defined that there must be space in at least two directions starting from the changed edge/node. Differences on the very left and very right of the graph were not automatically defined as having a lot of white space because the images of the graphs were of a certain size and in the study they were displayed on a colored background rather than a white one. Following this, in the shown example in Figure 4 difference A was defined having white space, while difference B and C were defined as having no white space.

The factor *color* describes whether the edges and nodes surrounding the difference were uniformly colored or multicolored. In Figure 5 we see an example. The edges and nodes in the near surrounding are all uniformly yellow, in the mutated graph one of the yellow edges turns into a green one.

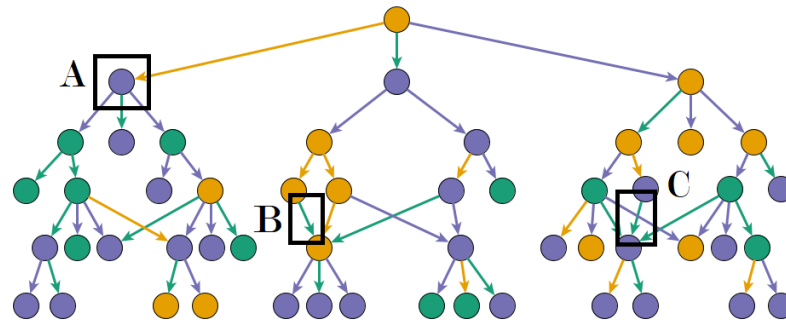


Figure 4: Example of defining white space.

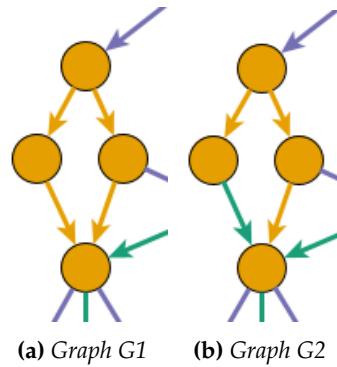


Figure 5: Shown here is a difference between graphs G1 and G2 within an uniform color environment.

To define whether the *edge length* is long or short, we did not have a specific measure, but specified it for each graph individually. We looked at the longest edges and the shortest edges to decide the categories.

4. Quantitative Results

We present results for the following six hypotheses:

- H1: Differences are more likely to be found in small graphs.
- H2: Differences are more likely to be found on the outside.
- H3: Differences are more likely to be found where white space is.
- H4: Differences are more likely to be found on the background of uniformly colored edges and nodes.
- H5: Differences are more likely to be found in changes to longer edges.
- H6: Differences with certain characteristics are found significantly earlier than others.

As can be seen in Table 2, hypotheses H1, H3, H4, and H5 were each confirmed by our results. Only hypothesis H2 showed no significant results, confirming the findings of Wallner et al. [165], and could therefore not be considered further. We calculated the Cramer's Value (see Table 2), which ranges from 0 to 1 and indicates how strong the perceived effects are. In our case, the effects are rather weak, but invite further investigation.

In addition, we computed GEE (generalized estimation equation). The outcome of this model supports the results of the single Chi-square analyses. The value $P(> |W|)$ is

Factor	p	χ^2	Cramer's V
Graph size	0.001972	9.58	0.064009409
Position (Shape)	0.1036	X	X
White space	2.8368E-05	17.524	0.086594324
Color	0.00016552	14.187	0.07791367
Edge length	0.01031426	6.58	0.145804929

Table 2: Results of the Chi-square test for quantitative data.

significant for size, white space, color and edge length, but not for position (see Figure 6). It can also be seen that no correlation could be found between the different factors and that they are independent of each other.

call:

```
geeglm(formula = found ~ size + position + whitespace + color +
edgelenh, family = binomial, data = GEETestdaten2,
id = interaction(graph, ID), corstr = "independence")
```

Coefficients:	Estimate	Std.err	Wald	Pr(> W)
---------------	----------	---------	------	----------

(Intercept)	0.08848	0.12160	0.529	0.46684	
size	-0.23655	0.08770	7.275	0.00699	**
position	0.00378	0.10355	0.001	0.97088	
whitespace	0.52713	0.11263	21.903	2.87e-06	***
color	0.47584	0.09919	23.013	1.61e-06	***
edgelenh	0.22208	0.07222	9.457	0.00210	**

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = independence

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	1.001	0.01559

Number of clusters: 2337 Maximum cluster size: 1

Figure 6: Resulting output from GEE.

As formulated in Hypothesis H6, we also analyzed the order in which the differences between the two graphs were perceived. Since the sequences vary in length (there are between two and five differences per graph), we decided to look more closely at the first and last differences found. We reasoned that the differences that participants found first would be the most salient, immediately noticeable, and would provide information about the importance of each factor that could be used to facilitate graph drawing. For each graph, we analyzed which differences were most often perceived as the first difference and which were most often perceived as the last difference. It can be seen that the most frequent first difference was changes in long edges, followed by a uniform surrounding as the second most frequent first difference and differences positioned inside the graph, while the most frequent last difference was white space, followed by short edges and differences positioned outside the graph. Table 3 shows the ranking of the most frequent first differences.

Type of Difference	Sum	First diff.	Percentage
Long edge	328	98	29.88
Uniformly colored	656	185	28.20
Position inside	1107	174	15.72

Table 3: Ranking of the most common first differences (Sum: Sum of found differences for 41 participants, First diff.: Found as first difference).

Type of Difference	[5]	[166]	[165]	Our Results
Shape	yes	yes	no	no
Edge Crossings	no	yes	yes	not tested
Density/White Space	no	yes	yes	yes
Graph Size	not tested	not tested	not tested	yes
Uniform Coloring	not tested	not tested	not tested	yes
Edge Length	not tested	not tested	not tested	yes

Table 4: Summary of what types of differences were found to have an impact on the comparison of participants.

To further support our descriptive statistics, we ran a binary logistic regression in SPSS. To summarize, it is more likely that a difference is perceived first when the difference has a uniform surrounding or when the edge is long. These values are comparable to our previous percentage calculations and confirm the statement.

After each graph pair, the participants were asked how difficult it was to find the differences and how certain they were that they have found all the differences. The average difficulty and certainty for every graph in descending order were afterwards analyzed. Graph B4G1 (Figure 7) that was considered the most difficult on average was also the graph where participants were most uncertain about finding the differences. Correspondingly, people felt most certain about finding the differences in the graph they considered easiest on average. For this part of the evaluation, we calculated the Spearman Correlation to determine the extent to which certainty, difficulty, and differences found were pairwise correlated. The Spearman correlations between certainty and difficulty, and between found differences and certainty were significant, while the correlation between found differences and difficulty was not significant. Wallner et al. [165] were able to calculate significant correlations between all three combinations in their analysis.

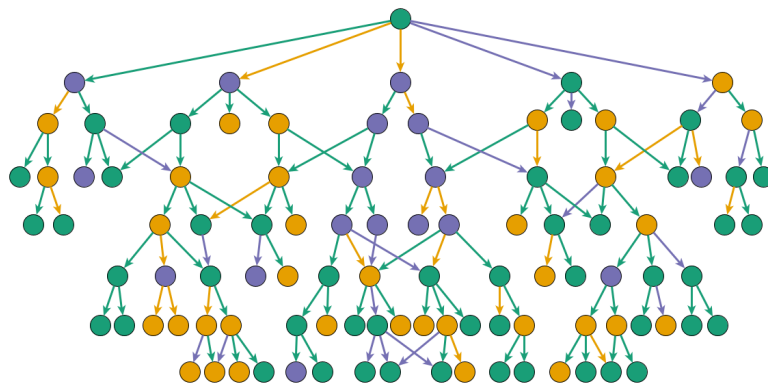


Figure 7: Graph B4G1 was on average considered the most difficult graph by the participants.

5. Discussion

Finally, we would like to put our results in the context of previous studies. In Table 4, we compare the main results of Wallner et al. and Ballweg et al. with our results presented here. Here it is easy to see the extent to which we have been able to confirm the results found so far while also giving an overview of our new results.

Our quantitative study showed that the factors white space, edge length, uniform/multicolored environment, and graph size were significantly impacting the perception of differences in the coloring of edges and nodes in DAGs under time constraints, while we did not obtain significant results for the aspect position (shape). Furthermore, no statistically significant interaction effect was found between the individual factors. Therefore, according to our hypotheses, we can now say that the differences are more likely to be found where white space is present, for changes with longer edges, for changes with a uniformly colored environment, and for small graphs. As a result, we think it is shown that coloring can have a significant impact on the perception and understanding of not only the graphs being displayed, but also their underlying data. In contrast, our hypothesis that differences are more likely to be found on the outside of a graph was rejected. Wallner et al. [166] observed similar results with respect to white space. In their study, white space was also cited as helpful for perceiving differences.

In the work of Wallner et al. [165], they also observed that the importance of external shape was not as great compared to the other factors. In our study, the external shape addressed by Wallner et al. could be compared to the "position" factor, where we define whether a difference is inside or outside a graph, which has an effect on the external envelope as perceived by the participants. Similar to their results, ours showed that this had no significant effect on perception.

The chi-squared test suggests that the size of the graph is a significant factor in the perception of differences, which can be compared with the results of Ballweg et al. [5] where the number of levels (depth) and the number of nodes on a given level most strongly influence the perception of similarity, therefore, our results also support these findings.

The factors most often perceived as the first difference were long edges and a uniformly colored environment. The uniformly colored environment made the differences in the graph more present, and these changes were often perceived first as a result, thus we suggest to highlight differences, that do not occur in such areas, to make it easier for the users to identify them.

What is interesting, in our opinion, is the lack of connection between a uniformly colored region and uniformly colored subgraphs within the structure. For example, the participants perceived nodes that were close to each other as one area, even though these nodes had no connection to each other or originated from different nodes. While the participants were focused on finding the differently colored edges and nodes, they seemed to ignore the underlying data structures of the DAGs.

In general the findings of this study suggest that the participants exhibited a strong inclination towards larger color areas when conducting visual comparisons. These larger color areas encompassed long edges, extensively colored regions spanning both nodes and edges, as well as substantial white surroundings in close proximity to the observed changes. The observed preference for such colored areas indicates a significant potential, particularly in the context of understanding the underlying graph data. These aspects have already been studied by Jianu et al. [75] and Saket et al. [133], but invite further investigation.

6. Future Work

This study has demonstrated the significant impact of changes in node and edge coloring between structurally identical node-link diagrams. Moving forward, it is crucial to investigate how individuals perceive these colored areas in more depth. Specifically, we aim to determine whether uniform coloring between parent-child nodes is necessary or if the visual context of a node or edge alone, irrespective of the connectivity of different parts, plays a more vital role in facilitating change detection. In addition, we are interested in exploring whether the uniform coloring of nodes alone is sufficient or if edges also need to be colored in the same manner. Furthermore, we are interested in exploring how different color schemes might yield distinct results, warranting further investigation.

The influence of different graph layouts on the visual comparison of colored DAGs remains an intriguing question that requires further exploration. Moreover, it is important to assess the impact of individually coloring edges and nodes on other types of graphs. Addressing these inquiries is our primary focus in future research endeavors, as they are essential for advancing scientific understanding in this field.

7. Acknowledgements

We thank the participants for taking part in our study.

Chapter 4

Conclusion and Future Directions

4.1 Conclusion on Human Perspectives on the Use of Color in Data Visualization for Quantitative Data

This work demonstrates that the perceptual discriminability of quantitative node colors in node-link diagrams is affected by the surrounding visual context. Across two studies, we found that a separating link significantly eases the detection of color differences between two nodes, regardless of its hue. Contrastingly a link directly connecting the nodes reduces discriminability. Complementary-colored links (or neutral gray links) amplify discriminability, leading us to recommend blue hues for nodes paired with yellow or gray links. In a second series of experiments, we showed that highlighting nodes with outward-directed, thin outlines have the least negative effect on color discrimination. For the color black, outlines show the least negative impact and when stronger emphasis is required, yellow or orange outlines outperform red outlines for blue nodes. Finally, the development and evaluation of the two new color schemes for large value ranges confirms that systematic hue selection combined with refined luminance gradients yields superior color discriminability, aesthetic appeal and uniform color spacing for data spanning large value ranges, compared to the original OMC scheme. Additionally an automation algorithm enables dynamic adjustment of the palette to support two to seven orders of magnitude compared to a fixed number of six in the previous color scheme. Together, these findings provide a coherent set of design guidelines for quantitative encoding to create more perceptually discriminable node-link diagrams (see [Figure 4.1](#)).

Recapitulating, we were able to answer the stated research questions. **RQ1.1:** *How should edge and node colors be chosen to ensure the discriminability between node colors that represent quantitative graph data in contrast to scatterplots?* When the links were colored in a complementary hue to the nodes, differences were perceived more clearly by the participants than when the link color was within the same hue as the nodes. In addition, the colors of the nodes were more discriminable when displayed in blue rather than yellow. **RQ1.2:** *How should node outlines be designed to ensure the discriminability between node colors that represent quantitative graph data in contrast to scatterplots?* The results indicate that for small to medium sized nodes the node colors are significantly more discriminable when the node outlines are directed outward. Black outlines have the least negative impact on the discriminability, while in cases of the need for a better highlighting effect, we advise to use yellow/orange outlines over red outlines for blue nodes. Furthermore, we found that thinner outlines improve the discriminability of the node colors, compared to thicker outlines, regardless of their color and direction. **RQ1.3:** *Which color scheme displays color differences for quantitative data with large value ranges in data visualization accurately for people with color vision deficiencies?* We present an improved and automated design of effective

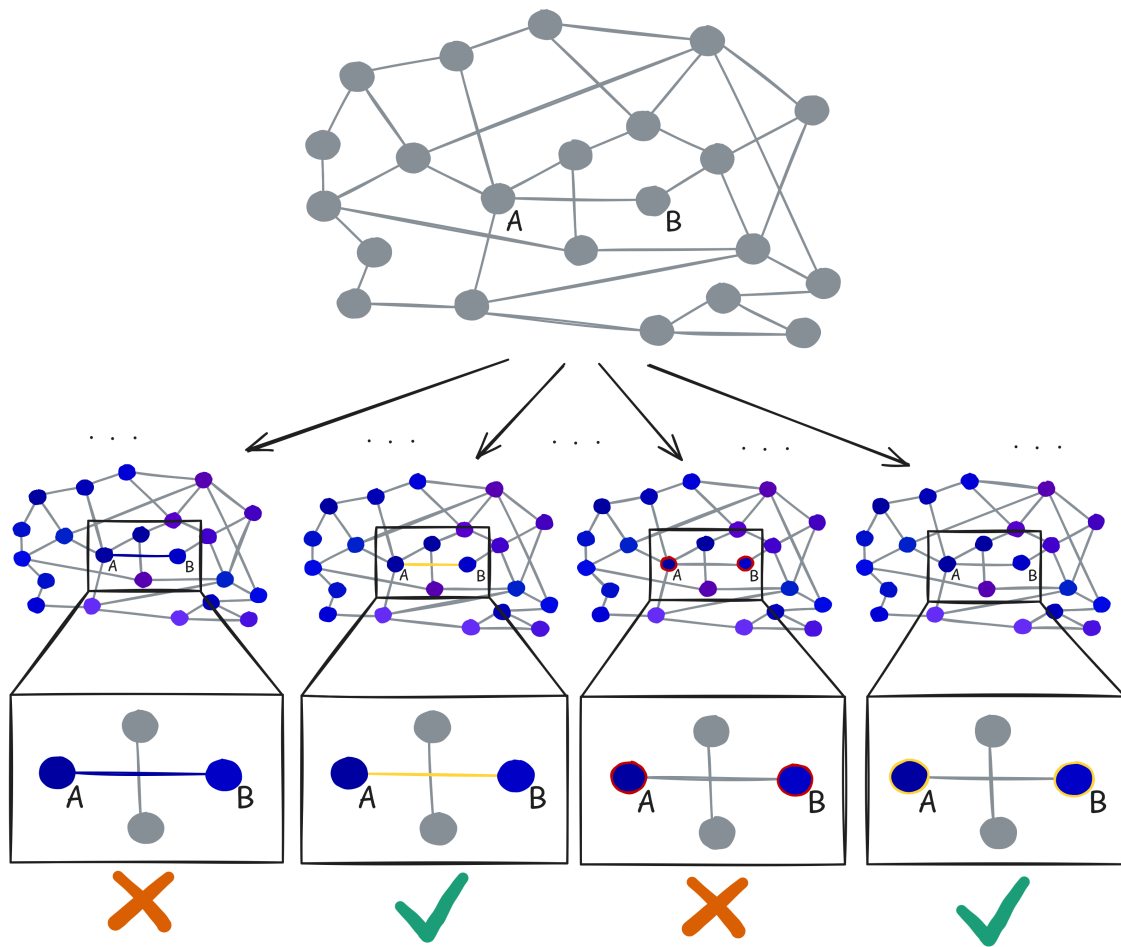


FIGURE 4.1: In this thesis we provide guidelines to support users in choosing suitable colors for creating visualizations to ensure the best possible color discriminability between nodes. In this example the color differences between nodes A and B are the same across all four node-link diagrams. However, depending on the edge and out-line colors, this color difference can be perceived differently.

color schemes for data with large value ranges. We ensure meeting a set of criteria-based design principles, including accessibility aspects for people with color vision deficiencies. Key improvements include the systematic selection of hues and refined luminance gradients for each magnitude.

The immediate **future work** could concentrate on extending the set of variables that may affect color discriminability of nodes in node-link diagrams. In particular, to examine how the size and shape of the nodes, the thickness of the edges, and the overall size of the graph influence the ability of users to distinguish colors. In addition, the impact of switching from undirected to directed node-link diagrams could be of interest, as the presence of arrowheads and flow direction may interfere with color perception and alter discriminable performance.

A longer-term goal is to carry out a comprehensive, systematic mapping of the color space that goes beyond the six hues examined in the previous studies. By exploring the full range of hue, saturation, and luminance values, we aim to derive general guidelines

that are applicable to any chosen node hue and color, thereby providing users with actionable recommendations for color selection.

Given the fast development of AI-driven visual analytics, it is also essential to explore color-related issues in that domain. Future research should therefore address two complementary aspects. First, to compare color discriminability for AI models (e.g., neural networks trained to recognize color patterns) with human perception, to understand any systematic divergences. Second, to investigate how the outputs of AI systems, visualized as trees, forests, or complex networks, can make use of color to highlight important data attribute values and thereby support the human interpretation of AI-generated visualizations.

4.2 Conclusion on Human Perspectives on the Use of Color in Data Visualization for Qualitative Data

The combined findings from our investigations reveal that both semantic associations and perceptual environmental factors critically shape how users distinguish color in node-link diagrams and scatterplots. In the domain of multi-label class data, we propose strategies to find suitable color schemes to encode the semantic meaning of each class label while simultaneously reflecting the relationships among classes that share identical labels. Our analysis indicates that the semantic association exerts a strong influence on the resulting palette, and the expert-driven use case demonstrates that applying our identified strategies can yield meaningful, task-relevant color mapping. These strategies include selecting a common color for the multi-label classes within the single-label classes and assigning the highest possible priority to it, using the average color values of the single-label classes to assign colors to the multi-label classes and choosing colors based on the environment of the classes. The algorithm we introduced executes these strategies. Following that, research on the perception of differences in colored node-link diagrams (specifically directed acyclic graphs) underscores the importance of visual context. Our experiment shows that four factors significantly increase participants' ability to detect color differences under time pressure: white space, edge length, the presence of a uniformly colored environment, and overall graph size. Taken together, these works provide important insights into the color design of node-link diagrams that use color to represent qualitative data attribute values in nodes.

Recapitulating, we were able to answer the stated research questions. **RQ2.1:** *What strategies should be used to create multi-label class color schemes with strong semantic associations?* Our research indicates that most individuals try to implement both the semantic associations to class labels and the connections between classes with identical labels when generating such color schemes. **RQ2.2:** *Which factors influence the perception of differences in categorically colored node-link diagrams?* Our quantitative study showed that the factors white space, edge length, uniform/multicolored environment, and graph size were significantly impacting the perception of differences in the coloring of edges and nodes in DAGs under time constraints.

A desired focus for **future work** on our algorithm for generating semantic color schemes for multi-label data is to extend it beyond the current limit of three label combinations. In addition, it would be an interesting progression to use AI systems to evaluate the semantic color associations that humans assign to different categories and, use these evaluations to automatically generate color palettes for any given set of data and the corresponding categories. Ultimately, this could lead to an AI-driven tool that recolors visualizations autonomously, taking the thematic context and the displayed categories into account.

A **future work** direction for graph comparison is to broaden the scope beyond directed acyclic graphs (DAGs) to include additional graph types, such as undirected, weighted, or hypergraph, and to evaluate a wider range of layout algorithms (e.g., force-directed, hierarchical, circular).

A further direction is to reproduce the already existing studies on the factors that influence graph comparability, this time using AI models. This line of research again comprises two complementary tracks: first, the investigation of which graph properties

(e.g., node position, whitespace, color uniformity) affect the ability of AI models to analyze and compare DAGs, and second to examine which of those same factors shape how humans compare AI-generated output, especially when the networks become large and complex.

Bibliography

- [1] Bodenfläche nach nutzungsarten und bundesländern. 2024.
- [2] J. Ahmad, E. Huynh, and F. Chevalier. When red means good, bad, or canada: Exploring people’s reasoning for choosing color palettes. In *2021 IEEE Visualization Conference (VIS)*, pp. 56–60, 2021. doi: [10.1109/VIS49827.2021.9623314](https://doi.org/10.1109/VIS49827.2021.9623314)
- [3] J. Albers. *Interaction of color / Josef Albers*. Yale University Press, New Haven [Connecticut], 50th anniversary edition, 4th edition. ed., 2013.
- [4] B. Alsallakh, L. Micallief, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. The state-of-the-art of set visualization. *Computer Graphics Forum*, 35(1):234–260, 2016. doi: [10.1111/cgf.12722](https://doi.org/10.1111/cgf.12722)
- [5] K. Ballweg, M. Pohl, G. Wallner, and T. Landesberger. Visual similarity perception of directed acyclic graphs: A study on influencing factors. vol. 22, p. 519–553, 09 2017. doi: [10.7155/jgaa.00467](https://doi.org/10.7155/jgaa.00467)
- [6] K. Batziakoudi, F. Cabric, S. Rey, and J.-D. Fekete. Lost in magnitudes: Exploring visualization designs for large value ranges. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI ’25, article no. 1170, 18 pages. Association for Computing Machinery, New York, NY, USA, 2025. doi: [10.1145/3706598.3713487](https://doi.org/10.1145/3706598.3713487)
- [7] K. Batziakoudi, S. Rey, and J.-D. Fekete. Beyond log scales: Toward cognitively informed bar charts for orders of magnitude values. *IEEE Transactions on Visualization and Computer Graphics*, 32(1):1120–1130, 2026. doi: [10.1109/TVCG.2025.3634795](https://doi.org/10.1109/TVCG.2025.3634795)
- [8] J. Bertin. *Semiology of graphics*. University of Wisconsin press, 1983.
- [9] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist, and J. Fekete. Graphdice: A system for exploring multivariate social networks. *Computer Graphics Forum*, 29(3):863–872, 2010. doi: [10.1111/j.1467-8659.2009.01687.x](https://doi.org/10.1111/j.1467-8659.2009.01687.x)
- [10] D. Bimler, J. Kirkland, and S. Pichler. Escher in color space: Individual-differences multidimensional scaling of color dissimilarities collected with a gestalt formation task. *Behavior Research Methods, Instruments, & Computers*, 36:69–76, 2004. doi: [10.3758/BF03195550](https://doi.org/10.3758/BF03195550)
- [11] J. Birch. Worldwide prevalence of red-green color deficiency. *J. Opt. Soc. Am. A*, 29(3):313–320, Mar 2012. doi: [10.1364/JOSAA.29.000313](https://doi.org/10.1364/JOSAA.29.000313)
- [12] C. Blair, X. Wang, and C. Perin. Quantifying emotional responses to immutable data characteristics and designer choices in data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1006–1016, 2024. doi: [10.1109/TVCG.2024.3456361](https://doi.org/10.1109/TVCG.2024.3456361)

- [13] A. Blake, G. Stapleton, P. Rodgers, and J. Howse. How should we use colour in euler diagrams? In *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*, VINCI '14, 10 pages, p. 149–158. Association for Computing Machinery, New York, NY, USA, 2014. doi: [10.1145/2636240.2636838](https://doi.org/10.1145/2636240.2636838)
- [14] P. Bogunovich and D. Salvucci. The effects of time constraints on user behavior for deferrable interruptions. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3123–3126. ACM, New York, NY, USA, 2011. doi: [10.1145/1978942.1979404](https://doi.org/10.1145/1978942.1979404)
- [15] D. J. Bora, A. K. Gupta, and F. A. Khan. Comparing the performance of l*a*b* and HSV color spaces with respect to color image segmentation. *CoRR*, abs/1506.01472, 2015. doi: [10.48550/arXiv.1506.01472](https://doi.org/10.48550/arXiv.1506.01472)
- [16] R. Borgo, J. Dearden, and M. W. Jones. Order of magnitude markers: An empirical study on large magnitude number detection. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2261–2270, 2014. doi: [10.1109/TVCG.2014.2346428](https://doi.org/10.1109/TVCG.2014.2346428)
- [17] D. Borland and R. M. T. Ii. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications*, 27(2):14–17, 2007. doi: [10.1109/MCG.2007.323435](https://doi.org/10.1109/MCG.2007.323435)
- [18] M. Bostock, V. Ogievetsky, and J. Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 9 pages, dec 2011. doi: [10.1109/TVCG.2011.185](https://doi.org/10.1109/TVCG.2011.185)
- [19] D. Braun, R. Borgo, M. Sondag, and T. von Landesberger. Reclaiming the horizon: Novel visualization designs for time-series data with large value ranges. *IEEE Transactions on Visualization & Computer Graphics*, 30(01):1161–1171, 2024. doi: [10.1109/TVCG.2023.3326576](https://doi.org/10.1109/TVCG.2023.3326576)
- [20] D. Braun, R. Borgo, M. Sondag, and T. von Landesberger. Design and evaluation of visualizations for large value ranges in multiple time series. *Information Visualization*, 25(1):66–84, 2026. doi: [10.1177/14738716251349501](https://doi.org/10.1177/14738716251349501)
- [21] D. Braun, K. Ebell, V. Schemann, L. Pelchmann, S. Crewell, R. Borgo, and T. von Landesberger. Color coding of large value ranges applied to meteorological data. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 125–129, 2022. doi: [10.1109/VIS54862.2022.00034](https://doi.org/10.1109/VIS54862.2022.00034)
- [22] H. C. Braun, K. Mukherjee, S. R. Gorelik, and K. B. Schloss. Affective color scales for colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 32(1):692–702, 2026. doi: [10.1109/TVCG.2025.3634775](https://doi.org/10.1109/TVCG.2025.3634775)
- [23] C. Brewer. Color use guidelines for data representation. 08 1999.
- [24] C. A. Brewer, G. W. Hatchard, and M. A. Harrower. Colorbrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30(1):5–32, 2003. doi: [10.1559/152304003100010929](https://doi.org/10.1559/152304003100010929)
- [25] S. Bridgeman and R. Tamassia. A user study in similarity measures for graph drawing. *Journal of Graph Algorithms and Applications*, 6(3):225–254, 2002. doi: [10.7155/jgaa.00053](https://doi.org/10.7155/jgaa.00053)

- [26] R. Bujack, T. L. Turton, F. Samsel, C. Ware, D. H. Rogers, and J. Ahrens. The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):923–933, 2018. doi: [10.1109/TVCG.2017.2743978](https://doi.org/10.1109/TVCG.2017.2743978)
- [27] M. Burch, W. Huang, M. Wakefield, H. C. Purchase, D. Weiskopf, and J. Hua. The state of the art in empirical user evaluation of graph visualizations. *IEEE Access*, 9:4173–4198, 2021. doi: [10.1109/ACCESS.2020.3047616](https://doi.org/10.1109/ACCESS.2020.3047616)
- [28] N. Camgöz, C. Yener, and D. Güvenç. Effects of hue, saturation, and brightness on preference. *Color Research & Application*, 27(3):199–207, 2002. doi: [10.1002/col.10051](https://doi.org/10.1002/col.10051)
- [29] R. C. Carter and L. D. Silverstein. Size matters: Improved color-difference estimation for small visual targets. *Journal of the Society for Information Display*, 18(1):17–28, 2010. doi: [10.1889/JSID18.1.17](https://doi.org/10.1889/JSID18.1.17)
- [30] G. Charness, U. Gneezy, and M. A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, 81(1):1–8, 2012. doi: [10.1016/j.jebo.2011.08.009](https://doi.org/10.1016/j.jebo.2011.08.009)
- [31] J. Chen, W. Yang, Z. Jia, L. Xiao, and S. Liu. Dynamic color assignment for hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):338–348, 2025. doi: [10.1109/TVCG.2024.3456386](https://doi.org/10.1109/TVCG.2024.3456386)
- [32] S. Chen, N. Andrienko, G. Andrienko, J. Li, and X. Yuan. Co-bridges: Pair-wise visual connection and comparison for multi-item data streams. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1612–1622, 2021. doi: [10.1109/TVCG.2020.3030411](https://doi.org/10.1109/TVCG.2020.3030411)
- [33] M. E. Chevreul and C. Martel. *The principles of harmony and contrast of colours, and their applications to the arts*. London : Longman, Brown, Green and Longmans, 1854.
- [34] J. Chuang, D. Weiskopf, and T. Moller. Hue-preserving color blending. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1275–1282, 2009. doi: [10.1109/TVCG.2009.150](https://doi.org/10.1109/TVCG.2009.150)
- [35] W. S. Cleveland. A color-caused optical illusion on a statistical graph. *The American Statistician*, 37(2):101–105, 1983. doi: [10.1080/00031305.1983.10482720](https://doi.org/10.1080/00031305.1983.10482720)
- [36] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: [10.2307/2288400](https://doi.org/10.2307/2288400)
- [37] A. M. Colman. Staircase method, 2009. doi: [10.1093/acref/9780199534067.013.7931](https://doi.org/10.1093/acref/9780199534067.013.7931)
- [38] A. M. Disrinama, M. Y. Santoso, G. Anindita, E. Setiawan, and J. Endrasmono. The effect of blue light filter on color perception in mobile device users. *AIP Conference Proceedings*, 2927(1):040013, 03 2024. doi: [10.1063/5.0194319](https://doi.org/10.1063/5.0194319)
- [39] P. Eades, S. Hong, A. Nguyen, and K. Klein. Shape-based quality metrics for large graph visualization. *Journal of Graph Algorithms and Applications*, 21(1):29–53, 2017. doi: [10.7155/jgaa.00405](https://doi.org/10.7155/jgaa.00405)
- [40] N. Earth. 1:110m cultural vectors, 2025. Last accessed: 2025-05-30.

- [41] V. Ekroll and F. Faul. New laws of simultaneous contrast? *Seeing and Perceiving*, 25(2):107–141, 2012. doi: [10.1163/187847612X626363](https://doi.org/10.1163/187847612X626363)
- [42] M. El-Assady, R. Kehlbeck, Y. Metz, U. Schlegel, R. Sevastjanova, F. Sperrle, and T. Spinner. Semantic color mapping: A pipeline for assigning meaningful colors to text. In *2022 IEEE 4th Workshop on Visualization Guidelines in Research, Design, and Education (VisGuides)*, pp. 16–22, 2022. doi: [10.1109/VisGuides57787.2022.00008](https://doi.org/10.1109/VisGuides57787.2022.00008)
- [43] M. D. Fairchild. Color appearance models. *John Wiley*, pp. 1–450, 2013. doi: [10.1002/9781118653128](https://doi.org/10.1002/9781118653128)
- [44] V. Filipov, A. Arleo, and S. Miksch. Are we there yet? a roadmap of network visualization from surveys to task taxonomies. *Computer Graphics Forum*, 42(6):e14794, 2023. doi: [10.1111/cgf.14794](https://doi.org/10.1111/cgf.14794)
- [45] Finnish Meteorological Institute. Cloudnet data portal. <https://cloudnet.fmi.fi/search/visualizations>. [Online; accessed 23-March-2023].
- [46] F. Franch, L. Nocciola, and A. Vouldis. Temporal networks in the analysis of financial contagion. pp. 1–52, 06 2022. doi: [10.13140/RG.2.2.25637.37607](https://doi.org/10.13140/RG.2.2.25637.37607)
- [47] A. Gaba, V. Setlur, A. Srinivasan, J. Hoffswell, and C. Xiong. Comparison conundrum and the chamber of visualizations: An exploration of how language influences visual design. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1211–1221, 2023. doi: [10.1109/TVCG.2022.3209456](https://doi.org/10.1109/TVCG.2022.3209456)
- [48] K. R. Gegenfurtner, D. Weiss, and M. Bloj. Color constancy in real-world settings. *Journal of Vision*, 24(2):12–12, 02 2024. doi: [10.1167/jov.24.2.12](https://doi.org/10.1167/jov.24.2.12)
- [49] G. A. Gescheider. *Psychophysics: The Fundamentals*. Psychology Press, 3rd ed., 1997. doi: [10.4324/9780203774458](https://doi.org/10.4324/9780203774458)
- [50] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. 10(4):289–309, 2011. doi: [10.1177/1473871611416549](https://doi.org/10.1177/1473871611416549)
- [51] E. E. GmbH. Eizo monitortest, 2025. Last accessed: 2025-05-20.
- [52] L. GmbH. Limesurvey: An open source survey tool.
- [53] I. M. Gołbiowska and A. Çöltekin. Rainbow dash: Intuitiveness, interpretability and memorability of the rainbow color scheme in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(7):2722–2733, 2022. doi: [10.1109/TVCG.2020.3035823](https://doi.org/10.1109/TVCG.2020.3035823)
- [54] C. C. Gramazio, D. H. Laidlaw, and K. B. Schloss. Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):521–530, 2017. doi: [10.1109/TVCG.2016.2598918](https://doi.org/10.1109/TVCG.2016.2598918)
- [55] C. Graniczowska, L. Pelchmann, T. v. Landesberger, and M. Pohl. Factors Influencing Visual Comparison of Colored Directed Acyclic Graphs. In M. Guthe and T. Grosch, eds., *Vision, Modeling, and Visualization*, pp. 107–114. The Eurographics Association, 2023. doi: [10.2312/vmv.20231232](https://doi.org/10.2312/vmv.20231232)

- [56] H. Hagh-Shenas, S. Kim, V. Interrante, and C. Healey. Weaving versus blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1270–1277, 2007. doi: [10.1109/TVCG.2007.70623](https://doi.org/10.1109/TVCG.2007.70623)
- [57] M. Harrower and C. A. Brewer. *ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps*, chap. 3.8, pp. 261–268. John Wiley & Sons, Ltd, 2011. doi: [10.1002/9780470979587.ch34](https://doi.org/10.1002/9780470979587.ch34)
- [58] S. He, W. Dong, J. Chen, J. Zhang, W. Lin, S. Yang, D. Xu, Y. Zhou, B. Miao, W. Wang, and F. Chen. Datacolor: unveiling biological data relationships through distinctive color mapping. *Horticulture Research*, 11(2):uhad273, 12 2023. doi: [10.1093/hr/uhad273](https://doi.org/10.1093/hr/uhad273)
- [59] C. Healey. Choosing effective colours for data visualization. In *Proceedings of Seventh Annual IEEE Visualization '96*, pp. 263–270, 1996. doi: [10.1109/VISUAL.1996.568118](https://doi.org/10.1109/VISUAL.1996.568118)
- [60] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2012. doi: [10.1109/TVCG.2011.127](https://doi.org/10.1109/TVCG.2011.127)
- [61] C. E. Helm and L. R. Tucker. Individual differences in the structure of color-perception. *The American Journal of Psychology*, 75(3):437–444, 1962. doi: [10.2307/1419867](https://doi.org/10.2307/1419867)
- [62] E. Hering. *Über die Beziehungen zwischen den Unterschieden der Lichtstärken der wirklichen Dinge und den tonfreien Helligkeitsunterschieden der Sehdinge*, pp. 62–100. Springer Berlin Heidelberg, Berlin, Heidelberg, 1920. doi: [10.1007/978-3-662-42443-8_3](https://doi.org/10.1007/978-3-662-42443-8_3)
- [63] M. Hertzum and K. D. Holmegaard. Perceived time as a measure of mental workload: Effects of time constraints and task success. *Int. Journal of Human-Computer Interaction*, 29(1):26–39, 2013. doi: [10.1080/10447318.2012.676538](https://doi.org/10.1080/10447318.2012.676538)
- [64] M. Hess, S. Bremm, S. Weissgraeber, K. Hamacher, M. Goesele, J. Wiemeyer, and T. von Landesberger. Visual exploration of parameter influence on phylogenetic trees. *IEEE Computer Graphics and Applications*, 34(2):48–56, 2014. doi: [10.1109/MCG.2014.2](https://doi.org/10.1109/MCG.2014.2)
- [65] M. Hlawatsch, F. Sadlo, M. Burch, and D. Weiskopf. Scale-stack bar charts. *Computer Graphics Forum*, 32(3pt2):181–190, 2013. doi: [10.1111/cgf.12105](https://doi.org/10.1111/cgf.12105)
- [66] M. Höhn, M. Wunderlich, K. Ballweg, and T. v. Landesberger. Width-scale bar charts for data with large value range. In *EuroVis 2020 - Short Papers*, pp. 103–107. Eurographics, 2020. doi: [10.2312/evs.20201056](https://doi.org/10.2312/evs.20201056)
- [67] R. Hu, Z. Ye, B. Chen, O. van Kaick, and H. Huang. Self-supervised color-concept association via image colorization. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):247–256, 2023. doi: [10.1109/TVCG.2022.3209481](https://doi.org/10.1109/TVCG.2022.3209481)
- [68] W. Huang, P. Eades, S.-H. Hong, and H. Been-Lirn Duh. Effects of curves on graph perception. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 199–203, 2016. doi: [10.1109/PACIFICVIS.2016.7465270](https://doi.org/10.1109/PACIFICVIS.2016.7465270)

- [69] W. Huang, S.-H. Hong, and P. Eades. Layout effects on sociogram perception. In P. Healy and N. S. Nikolov, eds., *Graph Drawing*, pp. 262–273. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. doi: [10.1007/11618058_24](https://doi.org/10.1007/11618058_24)
- [70] M. Höhn, M. Wunderlich, K. Ballweg, and T. v. Landesberger. Width-Scale Bar Charts for Data with Large Value Range. In A. Kerren, C. Garth, and G. E. Marai, eds., *EuroVis 2020 - Short Papers*. The Eurographics Association, 2020. doi: [10.2312/evs.20201056](https://doi.org/10.2312/evs.20201056)
- [71] ICoI. Cie recommendations on uniform color spaces, color-difference equations, and metric color terms. *Color Research & Application*, 2(1):5–6, 1977. doi: [10.1002/j.1520-6378.1977.tb00102.x](https://doi.org/10.1002/j.1520-6378.1977.tb00102.x)
- [72] A. Illingworth, R. Hogan, E. O’connor, D. Bouniol, M. Brooks, J. Delanoë, D. Donovan, J. Eastment, N. Gaussiat, J. Goddard, et al. Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, 88(6):883–898, 2007. doi: [10.1175/BAMS-88-6-883](https://doi.org/10.1175/BAMS-88-6-883)
- [73] S. Ishihara. *Tests for colour-blindness*. Kanehara Shuppan Company Japan, 1972.
- [74] J. Itten. *The elements of color; a treatise on the color system of Johannes Itten, based on his book The art of color*. A Basic color library. Van Nostrand Reinhold Co, New York, 1970.
- [75] R. Jianu, A. Rusu, Y. Hu, and D. Taggart. How to display group information on node-link diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 20(11):1530–1541, 2014. doi: [10.1109/TVCG.2014.2315995](https://doi.org/10.1109/TVCG.2014.2315995)
- [76] L. Joos, S. Jaeger-Honz, F. Schreiber, D. A. Keim, and K. Klein. Visual comparison of networks in vr. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3651–3661, 2022. doi: [10.1109/TVCG.2022.3203001](https://doi.org/10.1109/TVCG.2022.3203001)
- [77] B. Kale, M. Sun, and M. E. Papka. The state of the art in visualizing dynamic multivariate networks. *Computer Graphics Forum*, 42(3):471–490, 2023. doi: [10.1111/cgf.14856](https://doi.org/10.1111/cgf.14856)
- [78] R. Kehlbeck, J. Görtler, Y. Wang, and O. Deussen. Speuler: Semantics-preserving euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):433–442, 2022. doi: [10.1109/TVCG.2021.3114834](https://doi.org/10.1109/TVCG.2021.3114834)
- [79] H. A. Kestler, A. Müller, T. M. Gress, and M. Buchholz. Generalized Venn diagrams: a new method of visualizing complex genetic set relations. *Bioinformatics*, 21(8):1592–1595, 11 2004. doi: [10.1093/bioinformatics/bti169](https://doi.org/10.1093/bioinformatics/bti169)
- [80] R. G. Khueni. *Color: An introduction to practice and principles*. pp. 1–288, 2005. doi: [10.1002/0471687448](https://doi.org/10.1002/0471687448)
- [81] C. Kim, U. Hinrichs, S. M. Mohammad, and C. Collins. Lexichrome: Text construction and lexical discovery with word-color associations using interactive visualization. p. 477–488. Association for Computing Machinery, New York, NY, USA, 2020. doi: [10.1145/3357236.3395503](https://doi.org/10.1145/3357236.3395503)

- [82] Y. Kim, K. Thayer, G. S. Gorsky, and J. Heer. Color Names Across Languages: Salient Colors and Term Translation in Multilingual Color Naming Models. In J. Johansson, F. Sadlo, and G. E. Marai, eds., *EuroVis 2019 - Short Papers*. The Eurographics Association, 2019. doi: [10.2312/evs.20191166](https://doi.org/10.2312/evs.20191166)
- [83] F. A. A. Kingdom and N. Prins. *Psychophysics: A Practical Introduction*. Academic Press, London, 2nd ed., 2016.
- [84] D. Kringel, S. Malkusch, E. Kalso, and J. Lötsch. Computational functional genomics-based ampliseq™ panel for next-generation sequencing of 3 key genes of pain 4. *International Journal of Molecular Sciences*, 22:878, 01 2021. doi: [10.3390/ijms22020878](https://doi.org/10.3390/ijms22020878)
- [85] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. BELIV '06, 5 pages, p. 1–5. Association for Computing Machinery, New York, NY, USA, 2006. doi: [10.1145/1168149.1168168](https://doi.org/10.1145/1168149.1168168)
- [86] S. Lee, M. Sips, and H.-P. Seidel. Perceptually driven visibility optimization for categorical data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1746–1757, 2013. doi: [10.1109/TVCG.2012.315](https://doi.org/10.1109/TVCG.2012.315)
- [87] O. Lenz, F. Keul, S. Bremm, K. Hamacher, and T. von Landesberger. Visual analysis of patterns in multiple amino acid mutation graphs. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 93–102, 2014. doi: [10.1109/VAST.2014.7042485](https://doi.org/10.1109/VAST.2014.7042485)
- [88] J. Leung and A. Cockburn. Design framework for interactive highlighting techniques. *Found. Trends Hum.-Comput. Interact.*, 14(2–3):96–271, 179 pages, Nov. 2021. doi: [10.1561/11000000084](https://doi.org/10.1561/11000000084)
- [89] H. Levkowitz and G. Herman. Color scales for image data. *IEEE Computer Graphics and Applications*, 12(1):72–80, 1992. doi: [10.1109/38.135886](https://doi.org/10.1109/38.135886)
- [90] J. Li, Y. Liu, and C. Wang. Evaluation of graph layout methods based on visual perception. p. 1–7. Association for Computing Machinery, New York, NY, USA, 2016. doi: [10.1145/3009977.3010070](https://doi.org/10.1145/3009977.3010070)
- [91] K. Li, J. Li, Y. Sun, C. Li, and C. Wang. Color assignment optimization for categorical data visualization with adjacent blocks. *J. Vis.*, 26(4):917–936, 20 pages, jan 2023. doi: [10.1007/s12650-022-00905-z](https://doi.org/10.1007/s12650-022-00905-z)
- [92] J. Liang and M. L. Huang. Highlighting in information visualization: A survey. In *2010 14th International Conference Information Visualisation*, pp. 79–85, 2010. doi: [10.1109/IV.2010.21](https://doi.org/10.1109/IV.2010.21)
- [93] S. Lin, J. Fortuna, C. Kulkarni, M. Stone, and J. Heer. Selecting semantically-resonant colors for data visualization. In *Computer Graphics Forum*, vol. 32, pp. 401–410. Wiley Online Library, 2013. doi: [10.1111/cgf.12127](https://doi.org/10.1111/cgf.12127)
- [94] S. J. Little, S. L. Kosakovsky Pond, C. M. Anderson, J. A. Young, J. O. Wertheim, S. R. Mehta, S. May, and D. M. Smith. Using hiv networks to inform real time prevention interventions. *PLOS ONE*, 9(6):1–8, 06 2014. doi: [10.1371/journal.pone.0098443](https://doi.org/10.1371/journal.pone.0098443)

- [95] Y. Liu and J. Heer. Somewhere over the rainbow: An empirical assessment of quantitative colormaps. pp. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: [10.1145/3173574.3174172](https://doi.org/10.1145/3173574.3174172)
- [96] K. Lu, K. Reda, O. Deussen, and Y. Wang. Interactive context-preserving color highlighting for multiclass scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, article no. 823, pp. 1–15. Association for Computing Machinery, New York, NY, USA, 2023. doi: [10.1145/3544548.3580734](https://doi.org/10.1145/3544548.3580734)
- [97] S. LYi, J. Jo, and J. Seo. Comparative layouts revisited: Design space, guidelines, and future directions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1525–1535, 2021. doi: [10.1109/TVCG.2020.3030419](https://doi.org/10.1109/TVCG.2020.3030419)
- [98] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012. doi: [10.1109/TVCG.2012.279](https://doi.org/10.1109/TVCG.2012.279)
- [99] N. A. Macmillan and C. D. Creelman. *Detection theory: A user's guide*. Psychology Press, 2nd ed., 2005. doi: [10.4324/9781410611147](https://doi.org/10.4324/9781410611147)
- [100] M. Mahy, L. Van Eycken, and A. Oosterlinck. Evaluation of uniform color spaces developed after the adoption of cielab and cieluv. *Color Research & Application*, 19(2):105–121, 1994.
- [101] K. Marriott, H. Purchase, M. Wybrow, and C. Goncu. Memorability of visual features in network diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2477–2485, 2012. doi: [10.1109/TVCG.2012.245](https://doi.org/10.1109/TVCG.2012.245)
- [102] L. McNabb, R. S. Laramée, and M. Wilson. When size matters: Towards evaluating perceivability of choropleths. In G. K. L. Tam and F. Vidal, eds., *Computer Graphics and Visual Computing (CGVC)*, pp. 163–171. The Eurographics Association, 2018. doi: [10.2312/cgvc.20181221](https://doi.org/10.2312/cgvc.20181221)
- [103] A. Melamud, S. Hagstrom, and E. T. and. Color vision testing. *Ophthalmic Genetics*, 25(3):159–187, 2004. doi: [10.1080/13816810490498341](https://doi.org/10.1080/13816810490498341)
- [104] T. Mertz and J. Kohlhammer. Towards a quality approach to hierarchical color maps, 07 2024. doi: [10.48550/arXiv.2407.08287](https://doi.org/10.48550/arXiv.2407.08287)
- [105] W. Meulemans, N. H. Riche, B. Speckmann, B. Alper, and T. Dwyer. Kelpfusion: A hybrid set visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 19:1846–1858, 2013. doi: [10.1109/TVCG.2013.76](https://doi.org/10.1109/TVCG.2013.76)
- [106] S. Mittelstädt, D. Jäckle, F. Stoffel, and D. A. Keim. ColorCAT: Guided Design of Colormaps for Combined Analysis Tasks. In E. Bertini, J. Kennedy, and E. Puppo, eds., *Eurographics Conference on Visualization (EuroVis) - Short Papers*, pp. 115–119. The Eurographics Association, 2015. doi: [10.2312/eurovisshort.20151135](https://doi.org/10.2312/eurovisshort.20151135)
- [107] K. Mukherjee, B. Yin, B. E. Sherman, L. Lessard, and K. B. Schloss. Context matters: A theory of semantic discriminability for perceptual encoding systems. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):697–706, 2022. doi: [10.1109/TVCG.2021.3114780](https://doi.org/10.1109/TVCG.2021.3114780)

- [108] T. Munzner. Visualization analysis and design. pp. 1–2. A K Peters/CRC Press, 2014. doi: [10.1201/b17511](https://doi.org/10.1201/b17511)
- [109] P. Nardini, M. Chen, M. Böttinger, G. Scheuermann, and R. Bujack. Automatic improvement of continuous colormaps in euclidean colorspace. In *Computer Graphics Forum*, vol. 40, pp. 361–373. Wiley Online Library, 2021. doi: [10.1111/cgf.14313](https://doi.org/10.1111/cgf.14313)
- [110] C. Nothelfer, M. Gleicher, and S. Franconeri. Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception and Performance*, 43(9):1667–1676, 2017. doi: [10.1037/xhp0000314](https://doi.org/10.1037/xhp0000314)
- [111] J. R. Nuñez, C. R. Anderton, and R. S. Renslow. Optimizing colormaps with consideration for color vision deficiency to enable accurate interpretation of scientific data. *PloS one*, 13(7):e0199239, 2018. doi: [10.1371/journal.pone.0199239](https://doi.org/10.1371/journal.pone.0199239)
- [112] P. Paetzold, R. Kehlbeck, H. Strobel, Y. Xue, S. Storandt, and O. Deussen. Recteuler: Visualizing intersecting sets using rectangles. *Computer Graphics Forum*, 42(3):87–98, 2023. doi: [10.1111/cgf.14814](https://doi.org/10.1111/cgf.14814)
- [113] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. pp. 3659–3669. Association for Computing Machinery, New York, NY, USA, 2016. doi: [10.1145/2858036.2858155](https://doi.org/10.1145/2858036.2858155)
- [114] L. Pelchmann, R. Borgo, M. Pohl, and T. von Landesberger. Discriminability of node colors in node-link diagrams and the influence of link colors. *Visual Informatics*, 10(2):100288, 2026. doi: [10.1016/j.visinf.2025.100288](https://doi.org/10.1016/j.visinf.2025.100288)
- [115] L. Pelchmann, S. Bremm, K. Ebell, and T. v. Landesberger. Explorative Study on Semantically Resonant Colors for Combinations of Categories with Application to Meteorological Data. In C. Gillmann, M. Krone, and S. Lenti, eds., *EuroVis 2023 - Posters*. The Eurographics Association, 2023. doi: [10.2312/evp.20231061](https://doi.org/10.2312/evp.20231061)
- [116] L. Pelchmann, L. Jacobi, S. Bremm, and T. von Landesberger. Human strategies for semantic-based colors representing multi-label classes. In *ACM Symposium on Applied Perception 2026 (SAP '26)*, p. 10. ACM, New York, NY, USA, 2026. doi: doi.org/10.1145/3821409.3821477
- [117] A. Pister, C. Prieur, and J.-D. Fekete. ComBiNet: Visual Query and Comparison of Bipartite Multivariate Dynamic Social Networks. *Computer Graphics Forum*, pp. 290–304, Jan. 2023. doi: [10.1111/cgf.14731](https://doi.org/10.1111/cgf.14731)
- [118] S. M. Pizer. Intensity mappings to linearize display devices. *Computer Graphics and Image Processing*, 17(3):262–268, 1981. doi: [10.1016/0146-664X\(81\)90006-X](https://doi.org/10.1016/0146-664X(81)90006-X)
- [119] R. W. Pridmore. Complementary colors theory of color vision: Physiology, color mixture, color constancy and color perception. *Color Research & Application*, 36(6):394–412, 2011. doi: [10.1002/col.20611](https://doi.org/10.1002/col.20611)
- [120] H. C. Purchase, D. Carrington, and J.-A. Allder. Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering*, 7:233–255, 2002. doi: [10.1023/A:1016344215610](https://doi.org/10.1023/A:1016344215610)

- [121] H. C. Purchase, C. Pilcher, and B. Plimmer. Graph drawing aesthetics—created by users, not algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 18(1):81–92, 2010. doi: [10.1109/TVCG.2010.269](https://doi.org/10.1109/TVCG.2010.269)
- [122] G. Quadri and P. Rosen. Modeling the influence of visual density on cluster perception in scatterplots using topology, feb 2021. doi: [10.1109/TVCG.2020.3030365](https://doi.org/10.1109/TVCG.2020.3030365)
- [123] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafir. Do you see what i see? a qualitative study eliciting high-level visualization comprehension. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, article no. 204, 26 pages. Association for Computing Machinery, New York, NY, USA, 2024. doi: [10.1145/3613904.3642813](https://doi.org/10.1145/3613904.3642813)
- [124] C. Quijano-Chavez, B. Lee, N. Doerr, W. Buschel, M. Sedlmair, and D. Schmalstieg. Situated Brushing and Linking in Virtual and Augmented Reality . *IEEE Transactions on Visualization & Computer Graphics*, 32(05):4482–4491, May 2026. doi: [10.1109/TVCG.2026.3680714](https://doi.org/10.1109/TVCG.2026.3680714)
- [125] R. Rathore, Z. Leggon, L. Lessard, and K. B. Schloss. Estimating color-concept associations from image statistics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1226–1235, 2020. doi: [10.1109/TVCG.2019.2934536](https://doi.org/10.1109/TVCG.2019.2934536)
- [126] K. Reda, A. A. Salvi, J. Gray, and M. E. Papka. Color nameability predicts inference accuracy in spatial visualizations. *Computer Graphics Forum*, 40(3):49–60, 2021. doi: [10.1111/cgf.14288](https://doi.org/10.1111/cgf.14288)
- [127] K. Reda and D. A. Szafir. Rainbows revisited: Modeling effective colormap design for graphical inference. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1032–1042, 2021. doi: [10.1109/TVCG.2020.3030439](https://doi.org/10.1109/TVCG.2020.3030439)
- [128] D. Reimann, A. Schulz, N. Ram, and R. Gaschler. Color-encoded links improve homophily perception in node-link diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5593–5598, 2023. doi: [10.1109/TVCG.2022.3221014](https://doi.org/10.1109/TVCG.2022.3221014)
- [129] T.-M. Rhyne. Applying color theory to digital media and visualization (2nd ed.). pp. 1–227, 2024. doi: [10.1201/9781003500889](https://doi.org/10.1201/9781003500889)
- [130] N. H. Riche and T. Dwyer. Untangling euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1090–1099, 2010. doi: [10.1109/TVCG.2010.210](https://doi.org/10.1109/TVCG.2010.210)
- [131] A. R. Robertson. Historical development of cie recommended color difference equations. *Color Research & Application*, 15(3):167–170, 1990. doi: [10.1002/col.5080150308](https://doi.org/10.1002/col.5080150308)
- [132] B. E. Rogowitz, L. A. Treinish, and S. Bryson. How not to lie with visualization. *Computer in Physics*, 10(3):268–273, 05 1996. doi: [10.1063/1.4822401](https://doi.org/10.1063/1.4822401)
- [133] B. Saket, C. Scheidegger, and S. Kobourov. Comparing node-link and node-link-group visualizations from an enjoyment perspective. vol. 35, pp. 41–50, 2016. doi: [10.1111/cgf.12880](https://doi.org/10.1111/cgf.12880)
- [134] B. Saket, P. Simonetto, S. Kobourov, and K. Börner. Node, node-link, and node-link-group diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2231–2240, 2014. doi: [10.1109/TVCG.2014.2346422](https://doi.org/10.1109/TVCG.2014.2346422)

- [135] S. Salim, T. Pial, and K. Mueller. What is the color of serendipity? investigating the use of language models for semantically resonant color generation. *IEEE Transactions on Visualization and Computer Graphics*, 32(1):670–680, 2026. doi: [10.1109/TVCG.2025.3634243](https://doi.org/10.1109/TVCG.2025.3634243)
- [136] F. Samsel, T. L. Turton, P. Wolfram, and R. Bujack. Intuitive Colormaps for Environmental Visualization. In K. Rink, A. Middel, D. Zeckzer, and R. Bujack, eds., *Workshop on Visualisation in Environmental Sciences (EnvirVis)*, pp. 55–59. The Eurographics Association, 2017. doi: [10.2312/envirvis.20171105](https://doi.org/10.2312/envirvis.20171105)
- [137] A. Sarikaya and M. Gleicher. Scatterplots: Tasks, data, and designs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):402–412, 2018. doi: [10.1109/TVCG.2017.2744184](https://doi.org/10.1109/TVCG.2017.2744184)
- [138] M. Savary-Leblanc and X. L. Pallec. Interactive highlighting for digital uml class diagrams: a new feature. In *Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings, MODELS '22*, 10 pages, p. 247–256. Association for Computing Machinery, New York, NY, USA, 2022. doi: [10.1145/3550356.3561557](https://doi.org/10.1145/3550356.3561557)
- [139] V. Schemann and K. Ebell. Simulation of mixed-phase clouds with the icon large-eddy model in the complex arctic environment around ny-Ålesund. *Atmospheric Chemistry and Physics*, 20(1):475–485, 2020. doi: [10.5194/acp-20-475-2020](https://doi.org/10.5194/acp-20-475-2020)
- [140] K. B. Schloss. Color semantics in human cognition. *Current Directions in Psychological Science*, 33(1):58–67, 2024. doi: [10.1177/09637214231208189](https://doi.org/10.1177/09637214231208189)
- [141] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):810–819, 2019. doi: [10.1109/TVCG.2018.2865147](https://doi.org/10.1109/TVCG.2018.2865147)
- [142] K. B. Schloss, Z. Leggon, and L. Lessard. Semantic discriminability for visual communication. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1022–1031, 2021. doi: [10.1109/TVCG.2020.3030434](https://doi.org/10.1109/TVCG.2020.3030434)
- [143] P. Schulze-Wollgast, C. Tominski, and H. Schumann. Enhancing visual exploration by appropriate color coding. In *International Conference in Central Europe on Computer Graphics and Visualization*, pp. 203–210, 01 2005.
- [144] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005. doi: [10.1002/col.20070](https://doi.org/10.1002/col.20070)
- [145] D. Shi, A. Oulasvirta, T. Weinkauff, and N. Cao. Understanding and automating graphical annotations on animated scatterplots. In *2024 IEEE 17th Pacific Visualization Conference (PacificVis)*, pp. 212–221, 2024. doi: [10.1109/PacificVis60374.2024.00031](https://doi.org/10.1109/PacificVis60374.2024.00031)
- [146] S. Silva, J. Madeira, and B. S. Santos. There is more to color scales than meets the eye: A review on the use of color in visualization. In *2007 11th International Conference Information Visualization (IV '07)*, pp. 943–950, 2007. doi: [10.1109/IV.2007.113](https://doi.org/10.1109/IV.2007.113)

- [147] S. Silva, B. Sousa Santos, and J. Madeira. Using color in visualization: A survey. *Computers & Graphics*, 35(2):320–333, 2011. Virtual Reality in Brazil Visual Computing in Biology and Medicine Semantic 3D media and content Cultural Heritage. doi: [10.1016/j.cag.2010.11.015](https://doi.org/10.1016/j.cag.2010.11.015)
- [148] S. Smart and D. A. Szafir. Measuring the separability of shape, size, and color in scatterplots. pp. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: [10.1145/3290605.3300899](https://doi.org/10.1145/3290605.3300899)
- [149] U. Soni, Y. Lu, B. Hansen, H. C. Purchase, S. Kobourov, and R. Maciejewski. The perception of graph properties in graph layouts. vol. 37, pp. 169–181, 2018. doi: [10.1111/cgf.13410](https://doi.org/10.1111/cgf.13410)
- [150] M. Stone, D. A. Szafir, and V. Setlur. An engineering model for color difference as a function of size. *Color and Imaging Conference*, 22(1):253–253, 2014. doi: [10.2352/CIC.2014.22.1.art00045](https://doi.org/10.2352/CIC.2014.22.1.art00045)
- [151] O. Svenson and A. Maule, eds. *Time Pressure and Stress in Human Judgement and Decision Making*. Plenum Press, New York, 1993. doi: [10.1007/978-1-4757-6846-6](https://doi.org/10.1007/978-1-4757-6846-6)
- [152] D. A. Szafir. Modeling color difference for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):392–401, 2018. doi: [10.1109/TVCG.2017.2744359](https://doi.org/10.1109/TVCG.2017.2744359)
- [153] M. Tennekes and E. de Jonge. Tree colors: Color schemes for tree-structured data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2072–2081, 2014. doi: [10.1109/TVCG.2014.2346277](https://doi.org/10.1109/TVCG.2014.2346277)
- [154] M. Tennekes and M. J. H. Puts. cols4all: a Color Palette Analysis Tool. In T. Hoellt, W. Aigner, and B. Wang, eds., *EuroVis 2023 - Short Papers*, pp. 37–41. The Eurographics Association, 2023. doi: [10.2312/evs.20231040](https://doi.org/10.2312/evs.20231040)
- [155] K. M. Thyng, C. A. Greene, R. D. Hetland, H. M. Zimmerle, and S. F. DiMarco. True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, 29(3):9–13, 2016. doi: [10.5670/oceanog.2016.66](https://doi.org/10.5670/oceanog.2016.66)
- [156] C. Tominski, G. Fuchs, and H. Schumann. Task-driven color coding. In *2008 12th International Conference Information Visualisation*, pp. 373–380, 2008. doi: [10.1109/IV.2008.24](https://doi.org/10.1109/IV.2008.24)
- [157] B. E. Trumbo. A theory for coloring bivariate statistical maps. *The American Statistician*, 35(4):220–226, 1981. doi: [10.1080/00031305.1981.10479360](https://doi.org/10.1080/00031305.1981.10479360)
- [158] C. Tseng, G. J. Quadri, Z. Wang, and D. A. Szafir. Measuring categorical perception in color-coded scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, article no. 824, pp. 1–14. Association for Computing Machinery, New York, NY, USA, 2023. doi: [10.1145/3544548.3581416](https://doi.org/10.1145/3544548.3581416)
- [159] C. Tseng, A. Z. Wang, G. J. Quadri, and D. A. Szafir. Revisiting Categorical Color Perception in Scatterplots: Sequential, Diverging, and Categorical Palettes. In C. Tominski, M. Waldner, and B. Wang, eds., *EuroVis 2024 - Short Papers*. The Eurographics Association, 2024. doi: [10.2312/evs.20241073](https://doi.org/10.2312/evs.20241073)

- [160] C. Tseng, A. Z. Wang, G. J. Quadri, and D. A. Szafir. Shape it up: An empirically grounded approach for designing shape palettes. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):349–359, 2025. doi: [10.1109/TVCG.2024.3456385](https://doi.org/10.1109/TVCG.2024.3456385)
- [161] A. Tversky. Features of similarity. *Psychological review*, 84(4):327–352, 1977. doi: [10.1037/0033-295X.84.4.327](https://doi.org/10.1037/0033-295X.84.4.327)
- [162] J. Vihrovs., K. Prūsis., K. Freivalds., P. Ručevskis., and V. Krebs. An inverse distance-based potential field function for overlapping point set visualization. In *Proceedings of the 5th International Conference on Information Visualization Theory and Applications (VISIGRAPP 2014) - IVAPP*, pp. 29–38. INSTICC, SciTePress, 2014. doi: [10.5220/0004681100290038](https://doi.org/10.5220/0004681100290038)
- [163] T. von Landesberger, S. Diel, S. Bremm, and D. W. Fellner. Visual analysis of contagion in networks. *Information Visualization*, 14(2):93–110, 2015. doi: [10.1177/1473871613487087](https://doi.org/10.1177/1473871613487087)
- [164] M. Waldner, A. Karimov, and E. Gröller. Exploring visual prominence of multi-channel highlighting in visualizations. In *Proceedings of the 33rd Spring Conference on Computer Graphics, SCCG '17*, article no. 8, 10 pages. Association for Computing Machinery, New York, NY, USA, 2017. doi: [10.1145/3154353.3154369](https://doi.org/10.1145/3154353.3154369)
- [165] G. Wallner, M. Pohl, C. Graniczowska, K. Ballweg, and T. Landesberger. Influence of shape, density, and edge crossings on the perception of graph differences: An investigation under time constraints. pp. 348–356, 08 2020. doi: [10.1007/978-3-030-54249-8_27](https://doi.org/10.1007/978-3-030-54249-8_27)
- [166] G. Wallner, M. Pohl, T. von Landesberger, and K. Ballweg. Perception of differences in directed acyclic graphs: Influence factors & cognitive strategies. p. 57–64. Association for Computing Machinery, New York, NY, USA, 2019. doi: [10.1145/3335082.3335083](https://doi.org/10.1145/3335082.3335083)
- [167] Y. Wang, X. Chen, T. Ge, C. Bao, M. Sedlmair, C.-W. Fu, O. Deussen, and B. Chen. Optimizing color assignment for perception of class separability in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):820–829, 2019. doi: [10.1109/TVCG.2018.2864912](https://doi.org/10.1109/TVCG.2018.2864912)
- [168] C. Ware. *Information Visualization: Perception for Design*. Information Visualization: Perception for Design. Elsevier Science, 2013.
- [169] M. A. Webster and J. Mollon. The influence of contrast adaptation on color appearance. *Vision Research*, 34(15):1993–2020, 1994. doi: [10.1016/0042-6989\(94\)90028-0](https://doi.org/10.1016/0042-6989(94)90028-0)
- [170] J. Wu, J. Yang, M. Wei, K. Xiao, and S. Westland. Characterization of color differences for color palettes. *Color and Imaging Conference*, 2020:232–236, 11 2020. doi: [10.2352/issn.2169-2629.2020.28.37](https://doi.org/10.2352/issn.2169-2629.2020.28.37)
- [171] C. Xiong, V. Setlur, B. Bach, E. Koh, K. Lin, and S. Franconeri. Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):955–965, 2022. doi: [10.1109/TVCG.2021.3114823](https://doi.org/10.1109/TVCG.2021.3114823)
- [172] H. Yang, Y.-N. Li, and K. Zhang. Interactive influences of color attributes on color perception bias. *Vis. Comput.*, 36(5):925–937, 13 pages, may 2020. doi: [10.1007/s00371-019-01706-6](https://doi.org/10.1007/s00371-019-01706-6)

- [173] V. Yoghourdjian, D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu. Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Visual Informatics*, 2(4):264–282, 2018. doi: [10.1016/j.visinf.2018.12.006](https://doi.org/10.1016/j.visinf.2018.12.006)
- [174] L. Zhou and C. D. Hansen. A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(8):2051–2069, 2016. doi: [10.1109/TVCG.2015.2489649](https://doi.org/10.1109/TVCG.2015.2489649)
- [175] J. Zwinkels. *Light, Electromagnetic Spectrum*, pp. 1–8. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. doi: [10.1007/978-3-642-27851-8_204-1](https://doi.org/10.1007/978-3-642-27851-8_204-1)