Vision, Modeling, and Visualization (2023)
T. Grosch and M. Guthe (Eds.)

Factors Influencing Visual Comparison of Colored Directed Acyclic Graphs

C. Graniczkowska [1], L. Pelchmann[†2], T. von Landesberger[‡2] and M. Pohl[§1]

[1]TU Wien, Austria
[2]University of Cologne, Germany

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.

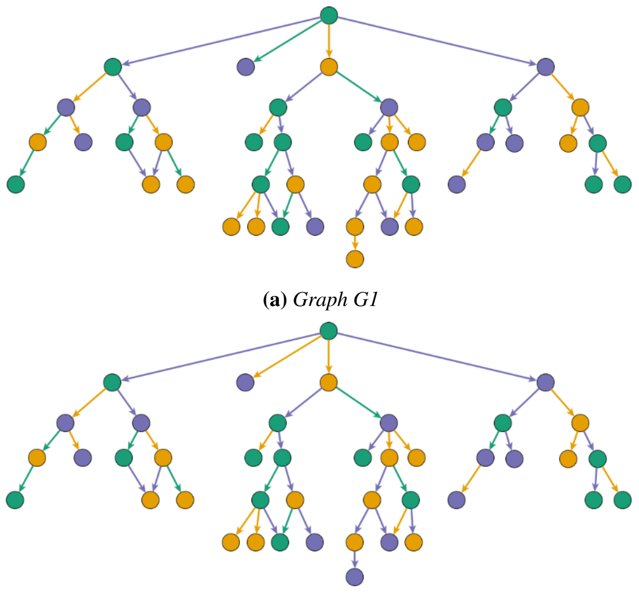
CCS Concepts

• ***Human-centered computing*** → ***Visualization;*** *Information visualization; Visualization design and evaluation methods;*

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. [vLDBF15a, LKB*14, HBW*14, EHNK17]. The comparison of these graphs relies heavily on the finding of differences, e.g. finding changes in nodes and edges, such ass 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. [FNV22, KMKL21] and require the perception of differences and similarities in the underlying data features [Tve77]. 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 [vLDBF15b, WPvLB19, GAW*11]. 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 [BPWL17, WPvLB19, WPG*20].



(a) *Graph G1*

(b) *Graph G2*

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.*

† pelchmann@cs.uni-koeln.de
‡ landesberger@cs.uni-koeln.de
§ margit@igw.tuwien.ac.at

DOI: 10.2312/vmv.20231232

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. [BPWL17, WPvLB19, WPG*20]. 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. [BPWL17] and Wallner et al. [WPG*20, WPvLB19].

Ballweg et al. [BPWL17] 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 [WPvLB19], 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 [WPG*20]. 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 [SM93]. Working under time constraints also increases the cognitive workload of participants [HH13, BS11]. 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 [WPvLB19, WPG*20] 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

2.2. Color

The usage of color in visualizations has been studied quite extensively. We may distinguish between two types of investigations,

Type of Dif.	**[BPWL17]**	**[WPvLB19]**	**[WPG*20]**
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.*

on the one hand investigations addressing how to adopt colormaps in visualizations to represent numerical values faithfully [LH18] [BTS*18] [RS21]. 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 ([Sza18]) 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 ([SS19]) 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 ([JRHT14]) investigated the influence of larger areas of color on the perception of groups of elements. Saket et al (2016) 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.

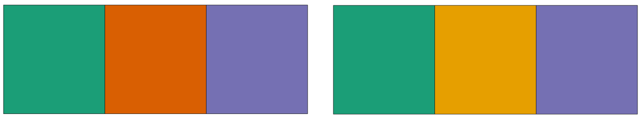
2.3. Graph Perception

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 [BHW*21]. To study the perception of visual properties of graphs, many works focus on single graphs. Lin et al. [LLW16] 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. [MPWG12] 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. Yoghourdjian et al. [YOG18] 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. [SLH*18] 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 [QR21]. The effects of curves and edge crossings on graph perception have also been adressed, by Huang et al. [HEHBLD16, HHE06]. The question of how to support comparison processes of visualized data is a very active area of research. Gleicher et al. [GAW*11] 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 (overlaying 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. [CAA*21] developed a system to compare multi-item data streams. Joos et al. [JJHS*22] developed a system to support comparison processes in VR. Pister et al. [PPF] developed a system for querying social networks and comparing the resulting selections. Researchers have also addressed more general questions. L'Yi et al. [LJS21] 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. [GSS*23] 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. [XSB*22] 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. [PKF*16] 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. [BT02] 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. [BPWL17, WPvLB19, WPG*20] 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 investigated 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



(a) *Color scheme for participants with deficits in color vision*

(b) *Color scheme for participants without deficits in color vision.*

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

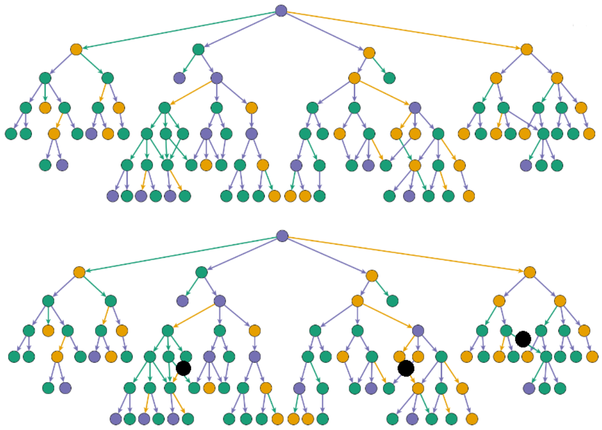


Figure 3: *Two exemplar large graphs shown to our participants. Differently colored nodes and links had to be marked in the bottom graph.*

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 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. [WPG*20]. 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

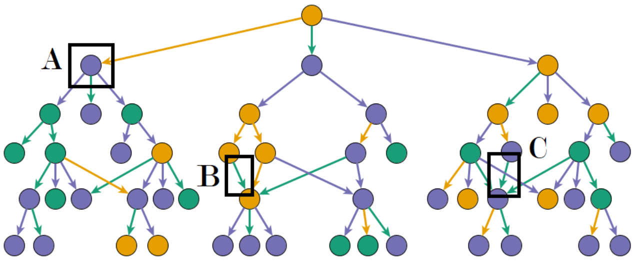


Figure 4: *Example of defining white space.*

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:

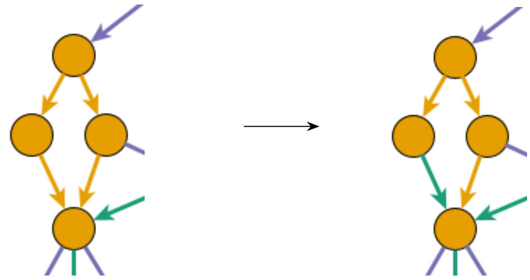
- graph size (small/large) (H1)
- position (not impacting the shape (inside)/impacting the shape (outside)) (H2)
- white space (yes/no) (H3)
- color (uniform/multicolored) (H4)
- edge length (no edge/short/long) (H5)

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.



(a) *Graph G1* **(b)** *Graph G2*

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

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.*

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.

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. [WPG*20], 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 significant for size,

Proceedings published by Eurographics - The European Association for Computer Graphics.

```
call:
geeglm(formula = found ~ size + position + whitespace + color +
edgelength, 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  ***
edgelength       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.*

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).*

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.

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.

To further support our descriptive statistics, we ran a binary logistic regression in SPSS. To summarize, it is more likely that a

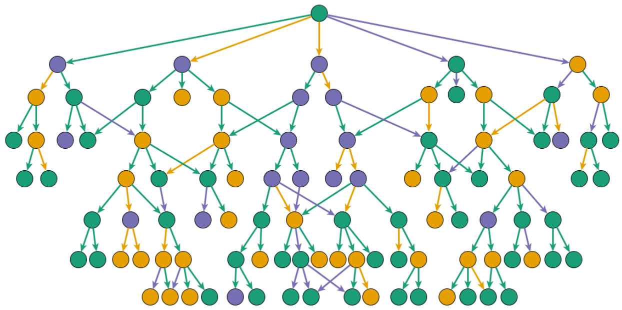


Figure 7: *Graph B4G1 was on average considered the most difficult graph by the participants.*

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. [WPG*20] were able to calculate significant correlations between all three combinations in their analysis.

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

Type of Difference	[BPWL17]	[WPvLB19]	[WPG*20]	Our Results
Shape	yes	yes	no	no
Edge Crossings	no	yes	yes	not investigated
Density/White Space	no	yes	yes	yes
Graph Size	not investigated	not investigated	not investigated	yes
Uniform Coloring	not investigated	not investigated	not investigated	yes
Edge Length	not investigated	not investigated	not investigated	yes

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

the outside of a graph was rejected.
Wallner et al. [WPvLB19] 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. [WPG*20], 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. [BPWL17] 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. [JRHT14] and Saket et al. [SSK16], 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.

Acknowledgements

We thank the participants for taking part in our study.

References

[BHW*21] BURCH M., HUANG W., WAKEFIELD M., PURCHASE H. C., WEISKOPF D., HUA J.: The state of the art in empirical user evaluation of graph visualizations. *IEEE Access 9* (2021), 4173–4198. doi:10.1109/ACCESS.2020.3047616. 3

[BPWL17] BALLWEG K., POHL M., WALLNER G., LANDESBERGER T.: Visual similarity perception of directed acyclic graphs: A study on influencing factors. vol. 22. doi:10.7155/jgaa.00467. 1, 2, 3, 7

[BS11] BOGUNOVICH P., SALVUCCI D.: The effects of time constraints on user behavior for deferrable interruptions. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), ACM, pp. 3123–3126. doi:10.1145/1978942.1979404.

[BT02] BRIDGEMAN S., TAMASSIA R.: A user study in similarity measures for graph drawing. *Journal of Graph Algorithms and Applications 6*, 3 (2002), 225–254. doi:10.7155/jgaa.00053. 3

[BTS*18] BUJACK R., TURTON T. L., SAMSEL F., WARE C., ROGERS D. H., AHRENS J.: 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 (2018), 923–933. doi:10.1109/TVCG.2017.2743978. 3

[CAA*21] Chen S., Andrienko N., Andrienko G., Li J., Yuan X.: Co-bridges: Pair-wise visual connection and comparison for multi-item data streams. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 1612–1622. doi:10.1109/TVCG.2020.3030411. 3

[EHNK17] Eades P., Hong S., Nguyen A., Klein K.: Shape-based quality metrics for large graph visualization. *Journal of Graph Algorithms and Applications 21*, 1 (2017), 29–53. doi:10.7155/jgaa.00405. 1

[FNV22] Franch F., Nocciola L., Vouldis A.: Temporal networks in the analysis of financial contagion. doi:10.13140/RG.2.2.25637.37607. 1

[GAW*11] Gleicher M., Albers D., Walker R., Jusufi I., Hansen C. D., Roberts J. C.: Visual comparison for information visualization. doi:10.1177/1473871611416549. 1, 3

[GSS*23] Gaba A., Setlur V., Srinivasan A., Hoffswell J., Xiong C.: Comparison conundrum and the chamber of visualizations: An exploration of how language influences visual design. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2023), 1211–1221. doi:10.1109/TVCG.2022.3209456. 3

[HBW*14] Hess M., Bremm S., Weissgraeber S., Hamacher K., Goesele M., Wiemeyer J., von Landesberger T.: Visual exploration of parameter influence on phylogenetic trees. *IEEE Computer Graphics and Applications 34*, 2 (2014), 48–56. doi:10.1109/MCG.2014.2. 1

[HEHBLD16] Huang W., Eades P., Hong S.-H., Been-Lirn Duh H.: Effects of curves on graph perception. In *2016 IEEE Pacific Visualization Symposium (PacificVis)* (2016), pp. 199–203. doi:10.1109/PACIFICVIS.2016.7465270. 3

[HH13] Hertzum M., Holmegaard K. D.: Perceived time as a measure of mental workload: Effects of time constraints and task success. *Int. Journal of Human-Computer Interaction 29*, 1 (2013), 26–39. doi:10.1080/10447318.2012.676538. 2

[HHE06] Huang W., Hong S.-H., Eades P.: Layout effects on sociogram perception. In *Graph Drawing* (Berlin, Heidelberg, 2006), Healy P., Nikolov N. S., (Eds.), Springer Berlin Heidelberg, pp. 262–273. doi:10.1007/11618058_24.

[JJHS*22] Joos L., Jaeger-Honz S., Schreiber F., Keim D. A., Klein K.: Visual comparison of networks in vr. *IEEE Transactions on Visualization and Computer Graphics 28*, 11 (2022), 3651–3661. doi:10.1109/TVCG.2022.3203001. 3

[JRHT14] Jianu R., Rusu A., Hu Y., Taggart D.: How to display group information on node-link diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics 20*, 11 (2014), 1530–1541. doi:10.1109/TVCG.2014.2315995. 3, 7

[KMKL21] Kringel D., Malkusch S., Kalso E., Lötsch J.: Computational functional genomics-based ampliseq™ panel for next-generation sequencing of 3 key genes of pain 4. *International Journal of Molecular Sciences 22* (01 2021). doi:10.3390/ijms22020878. 1

[LH18] Liu Y., Heer J.: Somewhere over the rainbow: An empirical assessment of quantitative colormaps. Association for Computing Machinery. doi:10.1145/3173574.3174172. 3

[LJS21] Lyi S., Jo J., Seo J.: Comparative layouts revisited: Design space, guidelines, and future directions. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 1525–1535. doi:10.1109/TVCG.2020.3030419. 3

[LKB*14] Lenz O., Keul F., Bremm S., Hamacher K., von Landesberger T.: Visual analysis of patterns in multiple amino acid mutation graphs. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 93–102. doi:10.1109/VAST.2014.7042485. 1

[LLW16] Li J., Liu Y., Wang C.: Evaluation of graph layout methods based on visual perception. Association for Computing Machinery. doi:10.1145/3009977.3010070. 3

[MPWG12] Marriott K., Purchase H., Wybrow M., Goncu C.: Memorability of visual features in network diagrams. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2477–2485. doi:10.1109/TVCG.2012.245. 3

[PKF*16] Pandey A. V., Krause J., Felix C., Boy J., Bertini E.: Towards understanding human similarity perception in the analysis of large sets of scatter plots. Association for Computing Machinery. doi:10.1145/2858036.2858155. 3

[PPF] Pister A., Prieur C., Fekete J.-D.: ComBiNet: Visual Query and Comparison of Bipartite Multivariate Dynamic Social Networks. 3

[QR21] Quadri G., Rosen P.: Modeling the influence of visual density on cluster perception in scatterplots using topology, feb 2021. doi:10.1109/TVCG.2020.3030365. 3

[RS21] Reda K., Szafir D. A.: Rainbows revisited: Modeling effective colormap design for graphical inference. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 1032–1042. doi:10.1109/TVCG.2020.3030439. 3

[SLH*18] Soni U., Lu Y., Hansen B., Purchase H. C., Kobourov S., Maciejewski R.: The perception of graph properties in graph layouts. *Computer Graphics Forum 37*, 3 (2018), 169–181. doi:https://doi.org/10.1111/cgf.13410.

[SM93] Svenson O., Maule A. (Eds.): *Time Pressure and Stress in Human Judgement and Decision Making*. Plenum Press, New York, 1993. doi:10.1007/978-1-4757-6846-6. 2

[SS19] Smart S., Szafir D. A.: Measuring the separability of shape, size, and color in scatterplots. Association for Computing Machinery. doi:10.1145/3290605.3300899. 3

[SSK16] Saket B., Scheidegger C., Kobourov S.: Comparing node-link and node-link-group visualizations from an enjoyment perspective. vol. 35, pp. 41–50. doi:https://doi.org/10.1111/cgf.12880. 7

[Sza18] Szafir D. A.: Modeling color difference for visualization design. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2018), 392–401. doi:10.1109/TVCG.2017.2744359. 3

[Tve77] Tversky A.: Features of similarity. *Psychological review 84*, 4 (1977), 327. doi:10.1037/0033-295X.84.4.327. 1

[vLDBF15a] von Landesberger T., Diel S., Bremm S., Fellner D. W.: Visual analysis of contagion in networks. *Information Visualization 14*, 2 (2015), 93–110. doi:10.1177/1473871613487087. 1

[vLDBF15b] von Landesberger T., Diel S., Bremm S., Fellner D. W.: Visual analysis of contagion in networks. *Information Visualization 14*, 2 (2015), 93–110. doi:10.1177/1473871613487087. 1

[WPG*20] Wallner G., Pohl M., Graniczkowska C., Ballweg K., Landesberger T.: Influence of shape, density, and edge crossings on the perception of graph differences: An investigation under time constraints. pp. 348–356. doi:10.1007/978-3-030-54249-8_27. 1, 2, 3, 4, 5, 6, 7

[WPvLB19] Wallner G., Pohl M., von Landesberger T., Ballweg K.: Perception of differences in directed acyclic graphs: Influence factors & cognitive strategies. Association for Computing Machinery. doi:10.1145/3335082.3335083. 1, 2, 3, 7

[XSB*22] Xiong C., Setlur V., Bach B., Koh E., Lin K., Franconeri S.: Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2022), 955–965. doi:10.1109/TVCG.2021.3114823. 3

[YOG18] Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Visual Informatics 2*, 4 (2018), 264–282. doi:https://doi.org/10.1016/j.visinf.2018.12.006. 3