Exploring the Effects of Colouring Graph Diagrams on People of Various Backgrounds

Beryl Plimmer[1], Ann Morrison[2] and Hendrik Knoche[2]

Dept Computer Science University of Auckland
beryl@cs.auckland.ac.nz
Department of Architecture, Design and Media Technology
{morrison|hk}@create.aau.dk

Abstract. Colour is one of the primary aesthetic elements of a visualization. It is often used successfully to encode information such as the importance of a particular part of the diagram or the relationship between two parts. Even so, there are few investigations into the human reading of colour coding on diagrams from the scientific community. In this paper we report on an experiment with graph diagrams comparing a black and white composition with two other colour treatments. We drew our subjects from engineering, art, visual design, physical education, tourism, psychology and social science disciplines. We found that colouring the nodes of interest reduced the time taken to find the shortest path between the two nodes for all subjects. Engineers, tourism and social scientists proved significantly faster with artist/designers just below the overall average speed. From this study, we contribute that adding particular colour treatments to diagrams increases legibility. In addition, preliminary work investigating colour treatments and schemes indicates potential for future gains.

Keywords: Colour encoding of diagrams, Graph diagram aesthetics, Subjects Disciplines' Effects

1 Introduction

The aesthetics of a diagram is acknowledged as an important factor in the ease to which a person can read and reason with the diagram. There has been considerable research into diagram layout and there are now well understood principles as to what layout strategies make a graph diagram easy (or hard) for people to interpret. These layout principles can be encoded into automatic layout algorithms so that the algorithms produce optimum layouts for human understanding.

There are many other elements of a diagram where aesthetic treatments can automatically be applied such as colour, fonts, line treatments, node shapes and sizes and iconic representations. There are bodies of knowledge about these aesthetic elements which could also inform diagram aesthetics. Attending to such aesthetics has been shown to be useful for user interface design [1, 2] reporting increased tolerance towards errors of 'beautiful' websites [2], and a strong correlation between the user's perception of a website's visual aesthetics and their perception of its credibility [3, 4] and trustworthiness [5].

Colour is a major component of aesthetics. It is often used in graph diagrams to encode information (e.g. Fig. 1) or simply to make the diagram look more pleasing. Yet, to our knowledge, there have been no experiments on the effects on human interpretation of different colour treatments of graph diagrams.

In this paper we report on an empirical shortest path experiment with three colour treatments applied to nodes of non-directional graphs. We selected people from a broad mix of disciplines as the subjects to investigate whether prior training or knowledge impacted on the experience. For example, we sought to determine if engineers would make the same errors, draw the same paths and/or prefer similar colour schemes as artist/designers, or people who specialise in physical education.

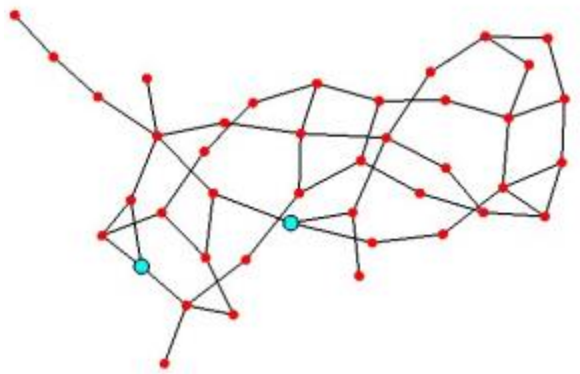


Fig. 1. Diagram used by [6] to explore effects of path straightness.

Important, as background for this study, we note that Gestalt theory and psychology argues that we visually identify features and complete shapes rather than seeing an ad-hoc assortment of simple lines, edges and bends as part of the laws of perceptual organization [7]. This 'seeing the whole' enables faster decision-making and uses less cognitive load by calling up information and understandings already gained as prior knowledge [8 ,9]. Furthering this work, Hatfield & Epstein [10] argue the perceptual system reduces its processing to the lowest possible effort by describing the outer world we encounter in the simplest means. Pomerantz [11], adds that colour perception acts in a similar holistic manner to shape perception as understood by Gestalt philosophy. This theory is well-placed to help us identify the role of colour in visual search tasks.

While Art and Design disciplines are clear in their use of colour [12], operating from a strong set of established design and colour guidelines that evolve as the formats change with time, artist/designers do not work from studies that 'prove' their theorems. Rather, they work from established conventions that derive from practice. While there are clearly successful results, this 'experience evidence' and the language that supports it, does not translate easily for more scientific disciplines who require contributions in the form of testing for evidence, findings and conclusions. There is considerable work to be done in the passing on of this knowledge from the art and design communities' perspectives to the diagrams community, with no easy way for direct translation of the accumulated knowledge.

In this paper we explore effects of both colour *treatments* and colour *schemes*. We use the term *treatments* as the semantic of the colour application – for example, colour to particular types of nodes. And colour *schemes* to be the colours applied – for example Fig. 1 the *treatment* is to colour the nodes for their shortest path task differently to all the other nodes on the diagram, the *scheme* is red/cyan.

The structure of this paper is as follows. In section 2 we review the work on diagram aesthetics and colouring of visualizations. Section 3 describes the experimental stimulus used including our design choices. Section 4 describes the methodology for the experiments followed by the analysis and results in Sections 5 and 6. Finally in Section 7 we discuss our findings, the implications of these findings and suggest future research.

2 Related work

The work of Purchase [13] on the aesthetics of diagram layout has been the basis of many investigations into the aesthetics of graph diagrams. Purchase showed that there are a number of layout principles that can be systematically applied to a diagram that make the diagram easier for people to understand. The most important of these principles is to decrease edge crossings. Purchase and others have expanded on these findings in various contexts and with dynamic graphs [e.g. 14, 15]. Purchase reported that the most significant layout aesthetics are to minimise edge crossings and edge bends and maximise symmetry. These findings have stood the test of time and are now regularly prioritised in layout algorithms [e.g. 16]. Without exception these studies have been conducted without regard to colour. Most studies have used black diagrams (nodes, edges and text) on a white background.

Colour is another fundamental aspect of aesthetics and visualizations. The perceptual and cognitive effects of colour have been studied for over 100 years. For more information we refer the reader to [17] that provides an excellent recent survey on the use of colour in visualizations and an overview of colour theory. It is notable that [17] has no human studies examples on the effectiveness of various colouring strategies from the field of abstract diagrams using graphs or set diagrams.

There have been numerous studies and algorithms proposed for automatic computer generation of colour schemes for particular contexts. These include [18] early work on contrasting and harmonious colour schemes for maps, and [19] for graph diagrams. These colouring generators are based on the theories of colour perception (see [20] as an example). However human experiments on the interpretation of the resultant colouring schemes are not reported. Others [21, 6] suggest that human experiments are necessary to validate such automated application of aesthetic elements, such as Ware's experiments for geographical map colouring [22].

Tsonos & Kouroupetroglou [23] investigate font size and styles and font/background colour combinations and found that in terms of colour combinations, black text on white background, white on blue, and green on yellow combinations were rated as the most pleasant.

Art and Design disciplines offer a variety of perspectives on the use of colour that fall beyond the scope of this study, however practical concepts such as colour relativity, intensity and temperature impact directly on the design considerations [24], with placement of colours to cause vibrations and the illusion of depth set up as background considerations.

Many abstract diagrams such as variants of graph diagrams and set diagrams use colour. There are a variety of ways colour is employed. Some use colour as a part of the syntax, for example [25, 6] use colour to differentiate nodes of interest. Others use colour as an extra level of information, for example [26] suggests the use of colour for alerting readers to important aspects of a diagram and [27] uses colour to indicate the degree of constraint satisfaction. In other cases, colour is used as eye-candy rather than for information for example [28]. None of these papers discuss the colour choices or reasons for them. So far as we are aware the only empirical studies into the effects of different colouring strategies are on concrete diagrams such as geospatial maps. We make an attempt with this initial study, to see if design understandings of colour can be translated into scientific evidence.

3 Experiment design

Clearly there is unlimited space for experimentation when applying colour to graph diagrams, with an infinite number of colours and colour combinations– at the extreme every node, edge and character could be a different colour. Choices must be made in order to produce empirical results. In this section we describe the experimental stimulus that we designed together with our reasoning for particular decisions.

Given that there is no prior work with abstract diagrams we chose to limit the experiment to fundamentals. We selected to experiment on fundamental graph diagrams – those with just nodes and edges. Based on Purchase's [13] graphs we devised three diagrams with equal numbers of nodes (16) and edges (28) and a similar distribution of edges per node, in addition each has an identifiable sub graph (see Fig. 2). We applied Purchase's [13] layout principles to the graphs so that there were no crossing edges, all edges were straight and that the diagrams were generally symmetric. This was a manual process as the layout algorithms in the software we used gave unsatisfactory results. We then flipped each diagram horizontally, vertically and horizontally and vertically resulting in 12 variants of the diagrams.

We then chose 3 different node pairs on each diagram for our shortest path questions. The paths varied in length from 2-5 edges. We visually identified these nodes in each diagram by allocating the Font Arial Rounded MT Bold, 18pt to the nodes that needed to be easily identified for answering the questions, whereas all other nodes were Helvetica Regular 10. We did this for all graphs regardless of colour treatments and schemes. The flipped graphs also had their label's font correspondingly flipped for easy legibility, that is, so the font was the correct way up.

We ran an informal experiment with 6 subjects and black & white rendering of the flipped diagrams to check whether people would recognize that there were only 3 diagrams. None of the subjects noticed this and all were surprised when we pointed it out to them.

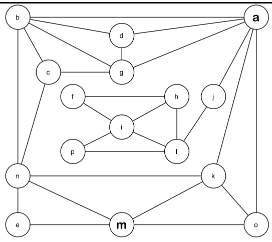
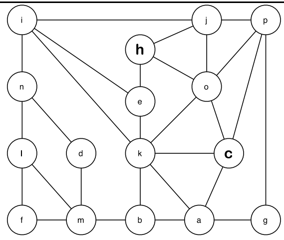
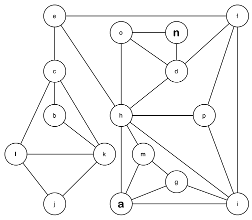
Graph 1	Graph 2	Graph 3
		
(cardinality 6x1,5x2,4x5,3x4,2x4)	(cardinality 6x1,5x1,4x6,3x5,2x5)	(cardinality 7x1,5x1,4x5,3x6,2x3)

Fig. 2. Three graphs used in experiment with black and white treatment. Nodes of interest have larger font and bolded labels. Cardinality refers to number of joining lines to one node. We refer to the black and white treatment as Set 1.

A major decision was on which colour treatments and schemes to apply. Black and white with the nodes of interest bolded as described above is retained as the default case. We considered a wide range of colouring treatments including: colouring all nodes the same, colouring nodes depending on their degree of connectiveness (cardinality), colouring nodes of interest as a highlighting strategy (design), colouring to highlight sub-regions differently (gestalt). From these we chose the two later treatments: nodes of interest (Set 2) and gestalt (Set 3).

The next decision was the colour schemes – that is what colours to use. We devised 3 colour schemes each consisting of two web safe colours. The colours we chose can be easily differentiated by the majority of the sighted public (taking into account various colour blindness conditions). The colours also largely cover the red-green-blue, red-yellow-blue primary colours spectrum. The colours for each graph were: Graph 1: D2FFFF (Pale Blue) and A5D5D5 (Grey_Green); Graph 2: FF0000 (Red) and FF6600 (Orange); and Graph 3: FF2425 (Scarlet) and A9ACD2 (Lilac) (see Fig. 3). A different colour scheme was applied to each graph.

We chose the colour combinations to create distinctions [12, 24] so information would sit on different levels, and to use primary (basic, pure), secondary (mix of two primaries) and tertiary (mix of primary and secondary) colours from intense saturated hues to luminance light-reflecting tints across the entire colour spectrum (designing around colour blindness). In the choices we made, we layered and separated out the nodes from the white background and the black text by using:

- Graph 1: Grey_Green/Pale Blue - cold, dull-hue, analogous (close to each other in the colour wheel) passive (visually receding) tints (colour added to white) with a high luminance (light reflective) value. We refer to this scheme as *calm.*
- Graph 2: Red/Orange – warm, active (visually appearing to advance to the foreground), highly-saturated (pure, undiluted by white), bright, intense analogous colours. We refer to this scheme as *bright.*

- Graph 3: Scarlet/Lilac - contrast between a warm, active, intense, saturated secondary colour (scarlet) and a cold, passive, dull, high luminance tertiary tint (lilac). We refer to this scheme as *contrast.*

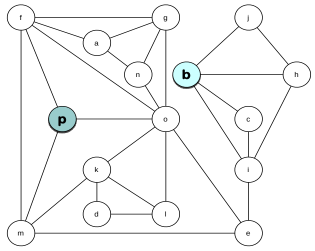
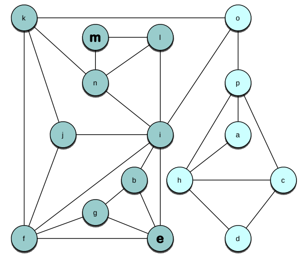
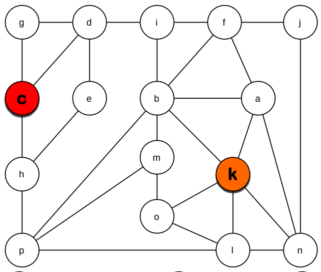
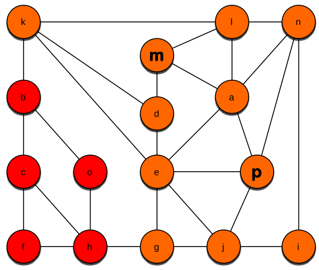
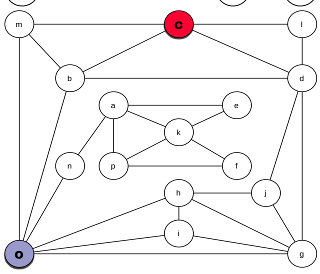
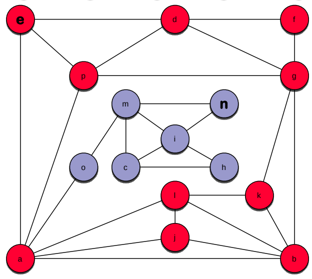
	Set 2 coloured nodes of interest	Set 3 all nodes coloured with subgraph differentiated
Graph 1 ***Calm***		
Graph 2 ***Bright***		
Graph 3 ***Contrast***		

Fig. 3. Two sets of colouring treatments with three colour schemes.

4 Methodology

The task selected was *shortest path* where the subjects were asked to mark the path that traversed the least number of edges between the two marked nodes. From the graph sets described in section 3 we randomly selected 9 graphs from each set (B&W, coloured nodes, coloured subgroups). For sets 2 and 3 the random selection was to randomly take 3 from each colour scheme. The selected graphs were added to a set of Powerpoint slides. Before these experimental graphs there were 3 training sections. The first was to familiarize the subjects with drawing on Powerpoint with a stylus. They were asked to draw a house and then on the following slide a person and any other object. This was followed by an explanation with examples of how to identify a shortest path. There were then 6 simple black and white graphs with a pair of bolded nodes that they completed with the facilitator. Following this they were told that the real experiment was beginning, however there were a further 6 training slides with increasingly complex graphs. The experimental graphs were then presented in the order black and white, coloured nodes and coloured subgroups. We chose not to counterbalance the sets to simplify the experiment, reasoning that with sufficient training (12 puzzle slides plus learning drawing with 3 slides before the experiment began), there should be no learning effect, and indeed the analysis of results indicated that this was true. In addition, based on pre-tests of fully randomized presentation in which the subjects found it difficult to identify the different colour schemes and even colour treatments, we created separate sets for each colour treatment, where for set 2 and 3 the colour schemes were in a fixed random order. Usually we would counterbalance the sets for order effects but it was more important that our subjects were able to reliably identify their preferences which they were not able to achieve in our randomized pilot study.

The subjects were asked to mark the shortest path on the Powerpoint slide and were given a maximum of 10 seconds per diagram to complete the task. The sessions were videoed and the time spent on each diagram and errors were later extracted from the video footage. After finishing all the tasks, the subjects completed a questionnaire on their perception of the effect of colouring the graphs and ended with a verbal debrief with the facilitator to elicit more qualitative information.

In selecting subjects we aimed for a range of disciplines from more visual (artistic) to more engineering styles of thinkers with a mix in between this polarized spectrum. To this end, we purposefully recruited a broad mix of 48 subjects. This mix was comprised of subjects with ages ranging from 20-38 and approximately even gender mix with 19 females and 29 males. Of these their background was: 18 artist/designers (8 female, 10 male); 8 male engineers; 8 mixed skill set of (2 more visually oriented females and 3 female and 3 males from engineering psychology); 6 physical education students (1 female and 5 male); and 8 tourism or social science students (5 female and 3 male). The individual sessions were facilitated by 28 students from a Masters User Experience Design class as part of their course work.

5 Analysis

From our 48 subjects, each solving 3x9 diagrams, we ended up with a total of 1296 solved diagrams. We investigated *completion time* and *errors* as dependent variables. *Completion time* was measured from the point at which the diagram appeared on the screen until the subjects had completed drawing the path from start to finish. This could include repairs in case a subject discovered having drawn a path that was too long and then drew another that was the shortest. Completion time was recorded with an accuracy of one second derived from the start and end time readings as identified and extracted from the video recordings. An *error* was scored when the subjects drew and settled on a path that was not the shortest.

Our independent variable was the

- three different *sets* (*black and white, coloured nodes* and *coloured subgroups*) and

For control variables we had

- the three different *diagrams* (*graphs 1, 2 and 3*) paired with three *colour schemes* (*calm, bright and contrast*) and black and white
- the number of times the subjects had seen the diagram before during the experiment (*diagram repetition*),
- the number of times the subjects had seen the diagram with the same start and end points (*puzzle repetition*),
- length of the solution counted in nodes between the start and the end node excluding these terminal nodes (*solution length*),
- the five different *disciplines* (*artist/designers, engineers, mixed skill set, physical education, tourism/social science).*
- age and
- gender.

Before introducing the control variables as predictors of *completion time* in a multiple regression we tested them individually. We checked for effects of *gender* (t-test), *discipline* (ANOVA) and *age* (single factor regression) on the subjects' averaged *completion time* and for effects of *diagrams*, *diagram repetition, puzzle repetition and solution length* (single factor regressions) on *completion time*. *Discipline* (F(4,43)=4.25, p<.01) and *solution length* (*b*=1.33, t(1)=19.5, p<.0001), were significant in predicting *completion time* as single factors and included in the subsequent analysis. We found no significant differences for *gender* t(1)=-0.78, p=.439, *age* t(1)=3.94, p=.763, *diagrams* t(2)=0.32, p=.725, *puzzle repetition* (*b*=0.35, t(1)=3.94, p=.763) or *diagram repetition* t(1)=0.6, p=.549 and excluded these from the subsequent multiple regression analysis.

We introduced *sets, demographics* and *solution length* as predictors for *completion time* into a multiple regression. Since we had this range of potential explanatory variables and could not refer to existing theory on which to base the model selection we carried out a stepwise linear regression. At each step the variable with the largest predictive power from the set with at least .05 significance was chosen and included in the model. Previously chosen predictors were removed from this set if their significance level went above a .10 cutoff.

In total our 48 subjects made 51 errors solving the diagrams. We used a logistic regression to analyze the independent and control variables as predictors for *error*.

6 Results

6.1 Completion Time

Controlling for all significant variables the regression showed that the subjects completed Set 2 with the coloured nodes of interest faster than the other two sets (*b*=-.32, t(2)=-3.60, p<.001) on average it took them 0.5 seconds less. *Solution length* significantly predicted *completion time* (b=1.29, t(1)=378.2, p<.0001) - the longer the solution path was the more time it took the subjects to complete the diagram (this is consistent with [6]). Our Set 2 had on average shorter solutions (2.9) than Set 1 (3) and Set 3 (3.4) but this was controlled for in the regression and the result that Set 2 was faster than Set 1 and 3 as described above still holds true.

Discipline (Fig. 4) was another variable that significantly predicted *completion time* (t(4)=31.6, p<.001). *Engineers* and *tourism/SocSci* were faster and both the *physical education* and *mixed skill set* were slower than the average (see Eq (1) below for the amounts of time according to the regression model). It should be noted that the data from subjects from each discipline were collected with the same procedure but different researchers facilitated the experiments.

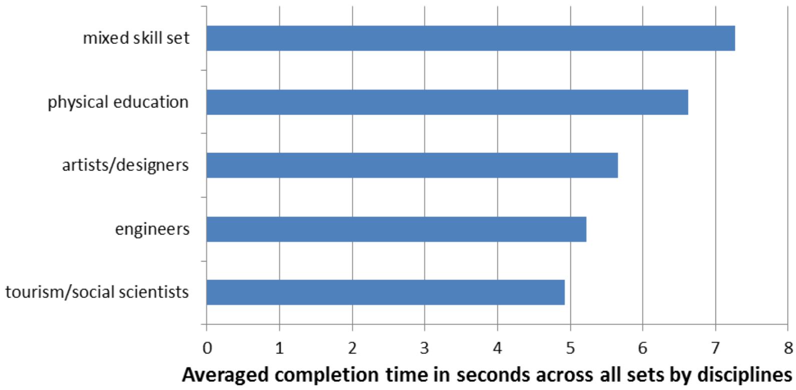


Fig. 4. Completion times by disciplines.

$$ct = 3.1 + 1.3 \times sl + \begin{cases} art, design: -0.3 \\ engineers: -0.7 \\ mixed\ skill\ set: 1.3 \\ physical\ educ: 0.7 \\ tourism, SocSci: -1 \end{cases} + \begin{cases} set1: +0.1 \\ set2: -0.3 \\ set3: +0.2 \end{cases} \qquad (1)$$

We tested for an interaction effect between *disciplines* and *sets* to see whether a given discipline benefited more or less from the different *sets*. The regression showed no evidence for an interaction indicating that our discipline groups were not differently affected by the sets. We also tested for an interaction effect of *solution length* and *sets* to check whether the subjects' performance for a given *solution length* was affected differently in the three *sets*. The regression yielded no significant effect.

The model to predict *completion time* (*ct*) in seconds of the experiment overall depicted in (1) was based on the three significant variables *solution length* (*sl*), *discipline* and *sets*. This model had a fit of R^2=.32, which is not very high but since our goal was not to model completion time but to understand whether it was significantly moderated by the predictor variables this was of no concern.

In order to be able to analyse the contribution of the different colour schemes without the confounding influence of the diagrams they appeared with, we needed to remove the diagrams as a factor. To this end we employed Set 1(black and white) as a baseline to establish the difficulty of the diagrams. We obtained an estimate of diagram difficulty by using a regression on the Set 1 data only with *solution length* and *disciplines* as variables alongside *diagrams*. All three variables were significant in predicting *completion time*. We then introduced the obtained values as a new variable *diagram difficulty* into our analysis and ran two regressions that compared Set 2 and Set 3 separately with Set 1 with *solution length, disciplines, diagram difficulty* and *colour scheme* as predictor variables. The first comparison (Set 1 vs. 2) looked at the advantage of colour schemes marking start and end nodes distinctly from the other nodes and the second (Set 1 vs. 3) tested how different colour schemes aided partitioning the Graph visually into sub-graphs. We did not include *sets* as a predictor since it could be expressed as linear combinations from *colour schemes*. In both regressions all the predictor variables that had been significant before (*solution length, disciplines,* and *diagram difficulty*) remained significant predictors.

In the Set 1 vs. 2 comparison we found a significant effect for colour combination (p=.004) as a whole. The brighter the intensity and the stronger the saturation of the colours between the background (white) and the colour scheme, the faster the subjects were: with black and white being significantly slower than all other colour schemes (*b*=.32, t(1)=2.91, p=.004), then calm (*b*=.22, t(1)=1.42, p=.15), contrast (*b*=-.27, t(1)=-1.92, p=.19) and bright – the fastest (*b*=-.28, t(1)=2.91, p=.13). However, the differences between the three colour schemes were not significant. For the second comparison of Set 3 and Set 1 we did not find a significant overall effect for *colour scheme* (F(3,847)=1.73, p=.16) despite two of the categorical levels being significant – specifically the bright condition being slowest (*b*=.39, t(1)=2.07, p=.039) and the contrast condition fastest (*b*=-.44, t(1)=-2.05, p=.04).

6.2 Error Rate

Whether or not our subjects made errors in the whole experiment was influenced by *gender, age* and *solution length* - they all contributed significantly to the variance of the data in the logistic regression. Errors were more likely to occur for women than for men ($\chi^2(1)=13.81$, p<.001), for younger than for older subjects ($\chi^2(1)=18.5$, p<.0001) and the longer the solution was ($\chi^2(1)=11.3$, p<.001). However, taken together these three variables only explained 13% of the variance of the data and we therefore do not report changes in odd ratios. Errors did not depend on the *diagram repetition, puzzle repetition, diagrams* or *sets*.

6.3 Self-Reported Results

The subjects were asked questions about the colour treatments and experience. They were asked to rank the colour treatments by preference (Fig. 5). The majority preferred Set 2 with Set 1 second favourite and just 5 subjects (4f, 1m) preferred Set 3. Interestingly, these 5 were significantly faster in finishing Set 3 than the other subjects and solved both Set 3 and Set 1 faster than Set 2. Secondly, the subjects ranked the colour treatments according to the confidence they had in completing a graph correctly in that condition. These results were almost identical to the preference ranking.

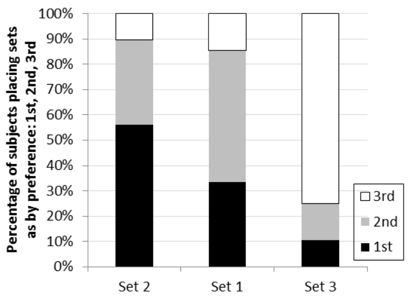


Fig. 5. Subjects rank preference for colour treatments.

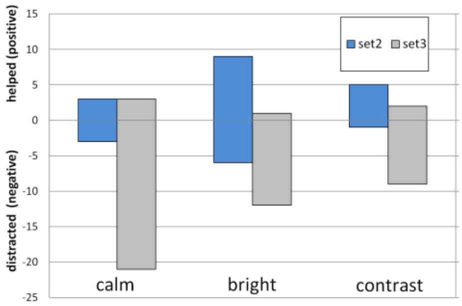


Fig. 6. Number of subjects commenting that the colour schemes helped or distracted in Set 2 and Set 3.

In the verbal debrief the subjects commented on their experiences solving the graphs in the different sets. We counted occurrences in which the subjects made references to the colour schemes being helpful in or whether they distracted from solving the graphs. The results are summarized in Fig. 6. For Set 2, the subjects reported that overall the colour schemes overall helped more than they distracted. This is in line with the performance measures. The subjects (9) most often mentioned the *bright* colour scheme as helpful, ahead of *contrast* (5) and *calm* (3). This tendency matches the performance results when comparing Set 1 to Set 2 where *bright* was fastest, *contrast* next, then *calm* with *black and white* being slowest. For Set 3 overall (where subjects were generally slower) the subjects reported the colours as being distracting - with *calm* (21) being worst. In contrast to Set 2 where *bright* was reported as more helpful ahead of *contrast*, in Set 3 more subjects (12) reported *bright* as being distracting than *contrast* (9).

7 Discussion and conclusions

Our goal with this work is to begin an investigation into the effect of different colouring strategies on graph drawing's legibility. This is an initial investigation to see if results 'understood' in the art and design community and, in the case of our Highlighted Nodes of Interest, often used in the diagrams community, can be proven and useful to the diagrams community. To this end we devised two colour treatments and three colour schemes which we then applied to graph diagrams. Participants from a range of disciplines completed shortest path tasks on the diagrams.

There are a number of limitations of the study and as pointed out in [29] these can affect the results. Clearly we could only choose a limited number of treatments and schemes. Our treatment selections are based on common strategies applied to graph diagrams and gestalt theories of visual grouping. The devised colour schemes are based on visual design practice and perception theory. The execution of the study also poses potential threats to validity. The facilitators were Masters students groups who worked with different discipline groups. While there were comprehensive instructions and training there is always the possibility of varying facilitator bias. Further we chose to fix the order of presentation of the diagrams. The diagrams were clustered into treatments, as pilot studies suggested that randomizing the treatments confused subjects. Likewise we found that fixing the order sets were presented in, assisted subjects to identify the colour treatments and to counter this we included sufficient pre-training examples to negate learning effects.

Of the colour treatments trialled, Set 2 with Highlighted Nodes of Interest, is an undeniable winner for the task tested. Subjects performed significantly better and most preferred it. There are many variations on this theme, such as Fig. 1, that could be further investigated. Certainly it would be straight forward to apply this treatment to interactive diagrams. It could be applied automatically to indicate, for example, faulty nodes in a network. Or as a response to user interaction, for example the user could colour a node of interest to mark it while searching the graph for other nodes.

Set 3, with the sub-graphs differentiated (gestalt), scored similarly to Set 1 (black & white) in performance but Set 1 was preferred by more. There was an interesting effect with 10% (5) subjects preferring Set 3 (gestalt) and preforming better with it: there is an obvious correlation between performance and preference. Why these subjects were different in preference and performance requires more data and investigation. The gestalt colour treatment enables subjects to more easily identify the connection paths between the sub-graphs. This set is interesting from another perspective. As the treatment does not directly relate to the task, the treatment could confound rather than assist subjects: however the visual grouping (seeing the whole gestalt) makes it easier to navigate.

Our colour schemes were based on design theory and visual perception. Using these theories, colour is used to effectively separate and layer information in a manner that results in intuitive correct interpretation of the data. We trialled three schemes each of which was designed to generate a layered effect on the 2-dimensional plane and create some kind of visual impact for the subjects. The calm colours pull the eyes into the background, we wanted to see if these colours would even be noticeable. The bright colours stood out from the background as we wanted to make sure people noticed there was colour. The contrast set created a 3-dimensionsal effect on the 2D plane with the soft cold colour drawing the eye back and the strong bright colour standing out from the background. While the number of data-points we have in each treatment/scheme is quite small, there are statistically significant results. When this is combined with subjects' comments on the effectiveness of different colour schemes it is a clear indication that colour choice matters. Interestingly, the bright colours that stood strongly out from the background were most helpful for Highlighted Nodes of Interest.

We have just scratched the surface of colour application to graph diagrams - we have applied two colour treatments and three schemes to only the nodes. Graph diagrams consist of nodes (that have borders and fill), edges which can also be coloured and have line treatments applied to them (heavier, lighter, dashed etc) and labels which have fonts, colours and weight.

There is an interesting dichotomy in the literature when one reviews the use of colours in graph interpretation studies. Most have been conducted on black and white diagrams similar to our Set 1. Those that have used colour, such as [25, 6] have done so without detailing the reasoning of their choices. Our results show that colour treatment and schemes effect subject performance, so careful consideration for colour choices should be taken. It would be interesting to test whether there are colour treatments that hinder interpretation. For example, if some nodes on a graph are highlighted with colour but these are *not* the nodes of interest for a particular problem, does this slow interpretation? Or do the coloured nodes provide spatial points of reference regardless?

Performance varied significantly in the different disciplines. However, we found colour treatments and schemes had a similar effect across the disciplines. Colours have cultural connotations, but whether these connotations affect the legibility of diagrams remains an open question.

Outside the diagrams community it is unusual to see a plain black and white diagram. General use, together with our results, suggest that colouring of the nodes of interest improves the legibility of the diagram for path-finding. Furthermore our model indicates a trend towards more efficiency with bright and/or contrasting colour schemes in comparsion to duller (calm) colours. Partitioning a diagram using colour, means the colour treatments heighten the underlying graph structure rather than highlighting the start and end points. The results for this treatment are mixed, suggesting further investigation is required. This first investigation into the effects of colouring graphs provides foundational evidence that colouring does effect legibility. More material from design knowledge needs to be translated and tested before a comprehensive set of colour guidelines could be defined and converted to algorithms for automatic application to diagrams.

References

1. Lindgaard, G., Fernandes, G., Dudek, C., and Brown, J.: Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression! Behaviour & Information Technology 25, 2, 115–126 (2006).
2. Hartmann, J., Sutcliffe, A., and Angeli, A. D.: Towards a Theory of User Judgment of Aesthetics and User Interface Quality. ACM Transactions on Computer-Human Interaction 15, 4, (2008).
3. Robins D., Holmes, J.: Aesthetics and credibility in web site design, Information Processing & Management. 44, 386-399 (2008)
4. Alsudani F., Casey, M.: The effect of aesthetics on web credibility, in BCS People and Computers. 512-519 (2009).
5. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P.: An Exploration of Relations Between Visual Appeal, Trustworthiness and Perceived Usability of Homepages. ACM Transactions on Computer-Human Interaction 18, 1, (2011)
6. Ware, C., Purchase, H., Colpoys, L., McGill, M.: Cognitive measurements of graph aesthetics, Information Visualization 1, 2, 103-110 (2002).
7. Köhler, W.: Gestalt Psychology. New York: Liveright. (1929)
8. Köhler, W.: Zum Problem der Regulation [On the problem of regulation]. In M. Henle (Ed.), The selected papers of Wolfgang Köhler (pp. 305–326). New York: Liveright. (1927)
9. Tufte, E.R.: Envisioning Information, 2nd Edition, Graphics Press USA, (1990)
10. Hatfield, G., & Epstein, W.: The status of minimum principle in the theoretical analysis of visual perception. Psychological Bulletin, 97, 20, 155–186 (1985).
11. Pomerantz, J. R.: Colour as a Gestalt: Pop out with basic features and with conjunctions, Visual Cognition. 14, 4-8, 619-628 (2006)
12. Albers, J.: Interaction of Color: 50th Anniversary Edition, Yale University Press. (2013)

13. Purchase, H.: Which aesthetic has the greatest effect on human understanding? In *Graph Drawing*, Springer Berlin Heidelberg, 248-261 (1997).
14. Purchase H. Samra, A.: Extremes Are Better: Investigating Mental Map Preservation in Dynamic Graphs, Diagrams Conference, G. Stapleton, J. Howse, and J. Lee, Eds. Springer-Verlag, LNAI 5223, 60-73 (2008)
15. Schneider E., Boucheix, J-M.: On Line Elaboration of a Mental Model During the Understanding of an Animation, in 4th International Diagrams Conference, D. Barker-Plummer, Cox, R. N. Swoboda, Eds. Springer-Verlag, LNAI 4045, 40-54, (2006).
16. Wybrow, M., Marriott, K. Stuckey, P.J.: Orthogonal Hyperedge Routing, Diagrams Conference. Springer-Verlag, LNAI 7352, 51-64, (2012).
17. Silva, S., Sousa Santos, B., Madeira, J.: Using color in visualization: A survey. Computers & Graphics 35, 2, 320-333, (2011).
18. Della Ventura, A., Schettini. R.: Computer aided color coding. Communicating with virtual worlds, Springer Japan, 62-75 (1993).
19. Dillencourt, M B., Eppstein D., Goodrich M.T..: "Choosing colors for geometric graphs via color space embeddings." In Graph Drawing, Springer Berlin Heidelberg, 294-305, (2007).
20. Ware, C.: Information visualization: perception for design. Elsevier, (2012).
21. Kosara, R., C. G. Healey, V. Interrante, D. H. Laidlaw, and C. Ware. "Visualization viewpoints." Computer Graphics and Applications, IEEE 23, no. 4 20-25 (2003).
22. Ware, C. "Color sequences for univariate maps: Theory, experiments and principles." Computer Graphics and Applications, IEEE 8, 5, 41-49, (1988).
23. Tsonos, D., Kouroupetroglou, G.: Modeling reader's emotional state response on document's typographic elements. Adv. in Hum.-Comp. Int., 2, January (2011)
24. Morten, J.L: Color Logic, Colorcom, (2008)
25. Burns, K., Bar-Gain Boxes: An Informative Illustration of the Pairing Problem, in Third International Conference, Diagrams, A. Blackwell, K. Marriott, and A. Shimojima, eds., Springer-Verlag: 379-381. (2004)
26. Lowe, R. Boucheix, M-J.: Supporting Relational Processing in Complex Animated Diagrams, in Diagrams, G. Stapleton, J. Howse, and J. Lee, eds., Springer-Verlag: Herrsching, Germany. 391-394. (2008)
27. Cheng, P.C.H. Barone, R.: Representing Rosters: Conceptual Integration Counteracts Visual Complexity, 3rd Conference, Diagrams, A. Blackwell, K. Marriott, and A. Shimojima, eds., Springer-Verlag: Cambridge, UK. 385-387. (2004)
28. Burns, K., Bayesian Boxes: A Colored Calculator for Picturing Posteriors, in 3rd International Conference, Diagrams, A. Blackwell, K. Marriott, and A. Shimojima, eds. Springer-Verlag: Cambridge, UK. 382-384, (2004).
29. Purchase, H.C.: Sketched Graph Drawing: A Lesson in Empirical Studies. Graph Drawing,. Springer International Publishing, 232-243 (2013).