TU WIEN Informatics

Einflussfaktoren in der Wahrnehmung von Unterschieden in gerichteten azyklischen Graphen

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Cynthia Graniczkowska, Bsc
Matrikelnummer 01427491

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof.in Mag.a rer.soc.oec. Dr.in phil. Margit Pohl,

Wien, 1. September 2022

Graniczkowska — Cynthia Graniczkowska

Margit Pohl — Margit Pohl

TU WIEN Informatics

Influence Factors in the Perception of Differences in Directed Acyclic Graphs

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Business Informatics

by

Cynthia Graniczkowska, Bsc
Registration Number 01427491

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof.in Mag.a rer.soc.oec. Dr.in phil. Margit Pohl,

Vienna, 1st September, 2022

Graniczkowska — Cynthia Graniczkowska

Margit Pohl — Margit Pohl

Acknowledgements

First of all, my gratitude belongs to Ao.Univ.Prof.in Mag.a rer.soc.oec. Dr.in phil. Margit Pohl, who supervised my thesis. I would like to express my sincere thanks for making this thesis possible, for all the input, suggestions, comments and feedbacks as well as for your precious time.

Furthermore I would like to thank all study-participants for their participation, because without them, these thesis would also not be possible.

And last but not least, I would like to thank my family for giving me the possibility to study and to my friends and my boyfriend for supporting me.

Kurzfassung

Der visuelle Vergleich von gerichteten azyklischen Graphen (DAGs) spielt eine wichtige Rolle in verschiedenen Anwendungsbereichen wie der Informatik, Biologie, Wirtschaft und vielen anderen. Mit dem ständigen Wachstum der Daten und der zunehmenden Komplexität in Graphen steigt der Bedarf an Vergleichswerkzeugen. Um herauszufinden, welche Faktoren die Wahrnehmung von Unterschieden in Graphen beeinflussen, muss Grundlagenforschung betrieben werden, um die Analyse von Graphen zu erleichtern. In dieser Masterarbeit werden Ergebnisse aus einer quantitativen und qualitativen Studie mit 49 Teilnehmern vorgestellt. Ziel der Studie war es, zu testen, ob Farbe und die Faktoren Graphengröße, Weißraum, Position, mehrfarbige/uniforme Umgebung und Kantenlänge die Wahrnehmung von Unterschieden zwischen Graphen und die Reihenfolge, in der sie wahrgenommen werden, beeinflussen, sowie die Suchstrategien der Teilnehmer zu ermitteln.

Abstract

The visual comparison of directed acyclic graphs (DAGs) plays an important role in various fields of application like in informatics, biology, economy and many more. With the constant growth of data and the increasing complexity in graphs, the demand for comparison tools is growing. To determine, which factors influence the perception of differences in graphs, there has to be done some basic research to facilitate analyzing graphs.

In this master thesis we will presents results from a quantitative and qualitative study with 49 participants. The goal of the study was to test whether color and the factors graph size, whitespace, position, multicolored/uniform surrounding, and edge length influence the perception of differences between graphs and the order in which they are perceived as well as to determine the search strategies of the participants.

Contents

CHAPTER 1

Problem Statement

The visual comparison of directed acyclic graphs (DAGs) plays an important role in various fields of application. It is, for example, needed in biology for the analysis of phylogenetic trees, gene regulation or for the prediction of infection diffusion as well as for example, in the financial sector for assessments of contagion in financial networks [30] [57].

Since visual comparison builds upon similarity judgements, similarity, and comparison are strongly connected [4]. As the amounts and therefore also the complexity of data is constantly growing, the demand for systems, which help with comparisons is increasing [20].

While there has been a lot of research in the readability of single graphs, the recognition and perception of differences in graphs and especially directed acyclic graphs are limited. Therefore, it is necessary to study which factors influence the perception of differences, to facilitate analyzing graph drawings [57].

CHAPTER 2

Related Work

2.1 DAG Definition

In graph theory and computer science a directed acyclic graph, also called "DAG", is a set of vertices and edges and has no directed cycles. Each edge has an orientation and connects one vertex to another. More formally a directed graph is defined as an ordered pair G= (V,E), where V stands for a finite, non-empty set of vertices and E stands for a set of edges that connect pairs of vertices[55] [47].

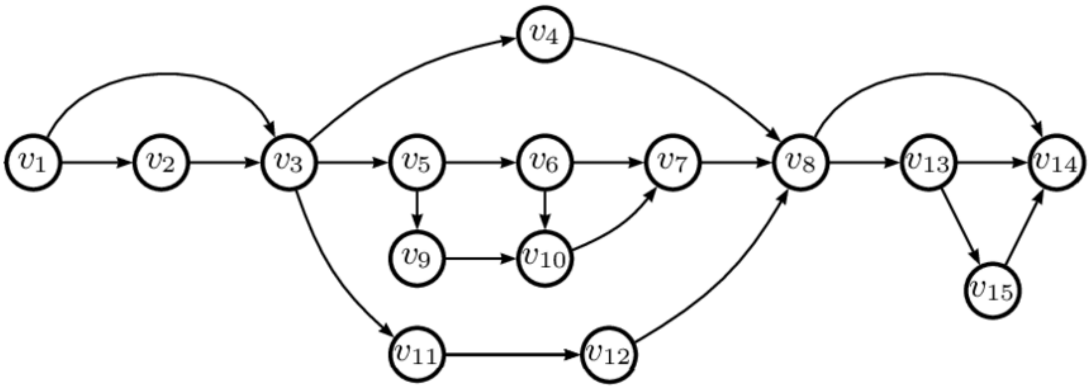

Figure 2.1: Example of a directed acyclic graph [17]

2.1.1 Topology

Every finite DAG has a topological sort, which means it has a list of all the vertices such that each vertex v appears earlier in the list than every other vertex reachable from v [32].

2.2 Usage of DAGs and graph comparison

The use of directed acyclic graphs is widespread and has a place in many different disciplines. Some examples of the usage and the comparison of graphs are mentioned in the following section.

2.2.1 Phylogenetic / Evolutionary trees

Phylogenetic trees are used to model the evolution of and relationships among sets of organisms. Those trees are not only interesting for evolutionary biologists, but are also relevant for pharmaceutical drug design. Researchers in this field have to understand the structural details of phylogenetic or evolutionary trees, which can be done by comparing different trees with each other [39] [38].

2.2.2 Citation Graphs

In a citation graph the nodes are documents, typically papers, journals, and books, including a publication date and the edges represent the citations from the bibliography of one document to another. The acyclic property of DAGs fits the needs of a citation graph, as there are no closed loops, because a paper cannot cite itself and in general two or more paper do not cite each other. Citation graphs enable finding publications, tracking the authorship of papers, the distribution of research findings, the possibility to enable communication between researchers to discuss and share the results of their work and they enable a computation of bibliometrics, which can display the venue or publication impact of a researcher in a specific research field [37] [7].

2.2.3 BPM

In the business process modeling field, for example, individual examples of workflows are recorded and are used to help and assist process engineers when creating and maintaining reference models for business processes. When a process engineer gets two different models of basically the same business process, he or she aims to visually compare them and visualize them into one single reference model [3].

2.2.4 Contagion in networks

Contagion is a process in which the collapse of one node leads to a chain reaction in a network, with neighboring nodes also collapsing. Such processes play an important role in different applications including gene regulation, infection diffusion prediction, supply chain management or financial network analysis. Analysts need visual comparison to detect and examine such contagion effects, and therefore the need to develop new visual analysis techniques is constantly increasing [30].

2.2.5 Detection of rumor spreading

Nowadays, when nearly everybody uses the internet and different social media platforms, people are able to gain information and spread information within seconds. On the one hand this kind of information can be useful, on the other hand it is often unreliable. Harmful rumors spread at extreme speed and can eventually cause people to panic. Therefore, mechanisms to detect such rumor information on the internet are of high importance. Directed acyclic graphs are also used to display rumor spreading. It starts with a source node, a single node, which knows about the rumor and has the intention to inform all other nodes. The size of the sharing tree is defined as the number of nodes (and hence the number of news sharers) and the height as the maximum path length from the root [24] [14].

2.2.6 Scheduling

Directed Acyclic graphs are also used for task scheduling. In such scheduling problems there are tasks and there is a set of constraints, which specify which task has to be completed before another one can start. This can be displayed in a DAG, were the nodes display the tasks and the edges display the direct prerequisite constraints. The aim is to find an order, where the tasks can be performed with respecting the constraints, which is also called topological sorting. If more than one task should be done at a time, parallel task scheduling is needed. To get the optimal parallel scheduling, walk relations in our DAGs have to be analyzed. Such scheduling problems appear for example if one efficiently wants to execute parallelized programs on a multiprocessor environment or for workflow scheduling [2] [28] [32].

2.2.7 DLT

Nowadays, distributed ledger technologies are used in many different fields of our everyday life including security, economic, juristic, and social aspects. Besides blockchains, which are the most popular technology, there are also other architectures of distributed ledger technologies. Hash graphs are seen as a very strong alternative to blockchain technologies. Hash graphs is a quite new technology based on graph architecture and especially on directed acyclic graphs. Blockchains can be interpreted as a simple directed acyclic graph where each block is seen as a node and has one predecessor, also called "parent" and one successor, also called "child". Just the first and last node are excluded to have both. The DAG of a hash graph displays the whole history of communication or also called "gossip" between the members of the population. Hash graph works with a gossip protocol and thus offers several advantages over other known DLTs [65].

2.3 Single Graphs

When it comes to the perception of visual features in graphs, most of the previous work concentrates on the perception of single graphs.

Li et al.[33] for example, explored the advantages and disadvantages of different graph layout methods as there are many different ones designed to meet different aesthetic criteria like same line length, evenly distributed points, or a minimum of edge crossings. As there is no layout which can fit all needs, it is important to select the best suitable one for the specific dataset. In their research work, they point out which nodes are visually more important, based on the degree of node, the number of neighbor nodes, the number of edges and the number of edge crossings in a certain surrounding area.

Mariott et al.[34] assessed the impact of layout features on the memorability of graphs including symmetry, collinearity, alignment, axis alignment and orthogonality. They conducted a study, were participants had to look at graphs and draw them from memory. As a result, they concluded that symmetry, collinearity, and orthogonality were significant features when recalling graphs, while node-alignment and parallel edges had no influence.

The work of Soni et al. [51] focuses on deciding which graph drawing layout helps a user best perceive a particular graph property, with the main focus on graph density and the average local clustering coefficient. They compared three different graph layout algorithms to decide which is best suitable for perceiving attributes.

2.4 Similarity

Similarity plays an important role in knowledge and behavior theories as it helps individuals to classify and form concepts and to make generalizations. In new approaches similarity is described as feature matching process and objects are collections of features [56].

Reisberg [45] on the other side points out, that similarity cannot be explained just as a comparison of features but is more complex. In his opinion, our theories about objects have an influence on similarity perception, for example if a tomato falls to the ground and is crushed, we still perceive it as a tomato, even though it no longer has the original shape.

Similarity is important for comparison because comparison builds upon similarity judgements. [4]. Perceiving similarities plays an important role in human cognitive processes, as similarities help us to categorize, organize and predict things in our world [21].

Pandey et al.[42] conducted an experiment to study the perception of similarity of scatterplots. Participants had to group plots according to their subjective similarity judgement. They identified key concepts of perceiving similarity.

The work from Bridgeman et al.[6] about similarity measures for graph drawing suggests that absolute and relative point positions are indeed important to the perception of similarity.

Ballweg et al.[4] tried to identify factors which influence the similarity perception of DAGs. They conducted a card-sorting study with 20 participants followed by a qualitative as well as a quantitative analysis to identify groups of DAGs which are perceived to be similar and the reasoning for the participant's choices. They created 69 small directed acyclic graphs with 6-9 nodes. The task was to group the graphs according to the perceived similarity. Participants also had to label the groups with the factors they used to create them and had to judge the difficulty of creating each group and the certainty of the group's consistency. The judgements were made on a Likert scale ranging from 1... very difficult/doubtful to 5... very easy/confident. Based on the work from Pandey et al.[42] they did a hierarchical clustering. For the clustering they determined properties like depth, symmetry, visual leaning, edge crossing, edge length, number of nodes per level, number of nodes with more than one parent node. Both, the quantitative and qualitative analysis show similar results and indicate that the number of levels, the number of nodes on a level and the overall shape of the graph mainly influence the similarity. Interestingly, edge crossing, which is defined to be an important factor in graph readability, seemed to not have a significant impact in the study.

2.5 Color

In visualizations, color is often used to encode values. To ensure, that those color encodings are effective, perceived differences in encoded values should match the differences in the underlying data. The problem is that most of the metrics for predicting perceived differences are designed under optimal conditions, while in practice environmental factors, display settings and properties of visualization design play a significant role. Many systems rely on conventional color difference metrics such as CIELAB, but this sometimes leads to underestimating color differences. It is in great demand, to provide an understanding of how to properly design visualizations when it comes to colors. Szafir constructed data-driven models for color difference perception in visualization, which focused on the mark types: points, bars, and lines [53].

Colormaps help to give more insight into data, as they can improve the efficiency and effectiveness of data perception. They play a key role in many different domains like computer vision, computer graphics, visualizations, and image processing. It is crucial for visualization designers to understand the techniques for colormap generation as well as general rules for choosing right colormaps for certain data [64].

2.6 Techniques for visual comparison

Gleicher et al. [20] Gleicher et al. believe that the development of future comparison tools will be facilitated if the understanding of comparison in general is developed. Although there are a lot of diverse systems and approaches, Gleicher et al. identified, that the basic types of techniques for visual comparison include: juxtaposition (showing different objects separately, superposition (overlaying objects in the same space), and explicit encoding.

2.7 Existing tools for comparison

In the last few years, a couple of systems have been introduced for the comparison of graphs, like for example systems comparing large phylogenetic trees, module relationships within software systems or genetic sequences. Even though those systems offer only limited help, as they are most commonly just prototypes, they show the value in developing tools which should support comparison tasks [20].
Some examples are listed and shortly described below:

2.7.1 SVG

Andrews et al. present a technique and prototype tool which support the comparison of graphs, the Semantic Graph Visualizer (SVG). It computes a merged graph and enables the analyst to visually compare the initial two graphs. The limitations are, that the input graphs have to be similar enough, otherwise the system does not work properly [3].

2.7.2 TreeJuxtaposer

The TreeJuxtaposer, proposed by Munzner et al. is a system for visually comparing hierarchies displayed as large trees with over hundred thousand nodes. The current limitations are that edge weights are not considered, although they are important, especially in the biological domain [39].

2.7.3 AlViz

The visual tool AlViz, proposed by Lanzenberger et al., which was introduced in the field of ontology alignment, uses clustering and side-by-side graph visualizations when comparing two ontologies [31].

2.8 Influence Factors and Strategies

Our study is strongly connected and build upon the previous research of Ballweg et al. [4], which was already mentioned in the section "Similarity" and the work from Wallner et al. [58] [57].

In a study with 40 participants, Wallner et al.[58] investigated if shape, density, and edge crossings have an influence on the perception of graphs and the order in which they are perceived. Similar research to Wallner et al. [57] was conducted, with the difference, that the study was done under time constraint, whereas the previous research was done without.
The dataset consisted of 16 graph pairs which were the same as in a previous work from Wallner et al. [57]. The graph pairs were displayed one above the other and the participants had to mark the perceived differences between those two graphs and afterwards answer questions rating the perceived difficulty and certainty of finding all differences.

Each difference was categorized based on the factors shape, which was concerned with the difference changing the outer hull i.e., the silhouette of the graph, the factor density, which was categorized in low, medium, and high and edge crossing which described if the change was a newly introduced edge crossing. According to the participants responses, with increasing graph complexity, the average certainty decreased, and the perceived difficulty increased.
A Spearman correlation calculation showed that the correlation between certainty and difficulty was significant as well as the correlation between found differences and difficulty and found differences and certainty.
They also found out, that shape had no significant influence on the perception. The results showed that an increase in density lowers the recognition of differences, and a new added edge crossing helps to recognize a change more easily. For comparing the sequences in which the differences were perceived, and as the graph pairs had different amounts of changes, they grouped them according to the number of differences and afterwards compared the significance of the factors.
Shape had an influence for graph pairs with two and four but not with three changes, lower density helped with finding differences in graph pairs with three and four changes while higher density was helpful for graphs with 2 changes. Edge crossing was helpful for recognizing differences in case of three and four changes, while it was the opposite for graphs with two changes.
So overall a change in the hull helped to spot a difference more quickly but was not significant when it comes to finding the difference at all. Density was an important factor when it came to spotting changes at all as well as a newly introduced edge crossing, which confirmed the previous work of Wallner et al. [57].
When graphs were perceived as being more difficult, also the uncertainty increased. Finding more differences, led the participants to increasing their perceived certainty.

Wallner et al. [57] conducted an exploratory qualitative study with 22 participants to determine which factors influence the perception of changes in directed acyclic graphs and which strategies were used to compare them using screen capturing and qualitative thinking aloud protocols.

The work was based on previous research from Ballweg et al [4] but differed from it as differences were investigated rather than similarities. Furthermore, pairwise comparison was used in this study, while Ballweg et al. used clustering. While in the work from Ballweg et al. the overall shape played an important role, the focus in this study laid on participants looking for individual differences. Moreover, this study concentrated on larger graphs.

A total of 16 graph pairs of different sizes and structures, based on 4 base graphs, were created, displayed side by side, and shown to the participants in a semi random order. The participants were asked to think aloud, and to mainly concentrate on the order in which they perceive the differences. The study was without any time constraint, as the main focus was on participants explanations. They conducted a qualitative content analysis with the verbalizations from the thinking aloud protocols. They identified eight factors which affect the perception of differences including edge crossings, shape, symmetry, white space, and density, number of layers, groups (of nodes) and arrowheads and calculated the number of verbalizations per category. The three most common used statements concerned symmetry, shape, and edge-crossing.

Symmetry was mentioned most often, as participants stated that it helped them recognizing differences. Participants used to look on the "opposite side" when having found a difference, hoping to find there another one. Shape was also of high importance for the participants, as they were looking for certain shapes and subparts of the graph which they compared with geometrical forms, as an aid for comparison. When it comes to edge crossing, the opinions of the participants were divided as some participants stated that edge crossing is making the process of comparing more difficult, while others were of the opinion that a newly introduced edge crossing is useful, especially if there was no edge crossing in the graph before but one in the altered graph. Participants felt that increased density, made it much more difficult to find differences, while a lower density helped them. On the contrary some stated that density helped them notice a new edge or node as the elements were so close together. Whitespace was found to be helpful for spotting changes in the graph pairs. Some participants used the number of the layers as an aid when comparing the graphs. What was stated as beneficial when looking for changes, were nodes which easily could be grouped. The arrowheads of the edges helped some participants to count the incoming edges, and on the other hand they helped spotting a new introduced edge as the arrowhead caused a spot to look denser. After analyzing the video recordings and thinking aloud protocols, they could distinguish three strategies on how the participants tried to assess the differences when comparing graphs, including "hierarchical", "layer-by-layer", and a mixture of both. The hierarchical approach included separating the graph into subtrees and traversing them using mainly a top to bottom approach, while a minority used a bottom-to-top approach. Some participants used an alternated top to bottom approach by going down a branch and coming back up the next branch. The subtrees were scanned from left to right. In the layer-by-layer approach the subjects looked for differences at the nodes and edges in one layer and continued with the next layer. Many participants sticked to just one strategy which they used for the whole study, while others adapted their strategy depended on the overall structure of the graph and its subparts.

CHAPTER 3

Research Methodology

In this master thesis we first conducted a literature review and afterwards used the explanatory mixed methods design that involved a quantitative data collection phase followed by a qualitative data collection.

3.1 Literature Review

Simply put, a literature review is a systematic way of collecting and summarizing prior research. A well-executed review as a research method creates a solid foundation for knowledge growth and theory development. It can provide an overview of areas in which research is inconsistent and interdisciplinary. Besides, it is perfectly suitable for summarizing research findings and revealing areas in which more detailed research is needed. Literature reviews are used to describe previous research, to evaluate the research area and to justify the research question and the hypotheses of the study.

However traditional ways of describing literature are often not done systematically and thorough enough. This can subsequently lead to authors making flawed assumptions and building their research on [48].

Over the years, many different types of literature reviews have emerged, but one can generally distinguish 4 main types: narrative, also called traditional review, systematic review, meta-analysis, and meta-synthesis.

3.1.1 Traditional or narrative literature review

The main purpose of this type of review is to analyze and summarize existing literature, by providing a comprehensive background, identifying research gaps, recognizing inconsistencies, and highlighting new research streams. It is best used for developing theoretical and conceptual frameworks as well as helping to determine or define research questions or hypotheses.

3.1.2 Systematic literature review

The purpose of a systematic literature review is to find all published and unpublished research work relating to the specified topic to ensure that no existing knowledge is missed. Systematic reviews use explicit and rigorous criteria to identify, critically evaluate and to synthesize and compare the literature relevant for the chosen topic.

3.1.3 Meta-analysis

When conducting a meta-analysis, one takes the chosen literature findings and analyze them by using standardized statistical procedures. This method helps to draw conclusions and to detect patterns and relationships.

3.1.4 Meta-synthesis

Meta-Synthesis is a non-statistical procedure, which evaluates and analyses findings from qualitative studies and its purpose is to build on previous interpretations and conceptualizations [40] [13].

For our work, we will be using a **narrative literature review**, as it best suits our needs.

3.2 Quantitative research

Quantitative methods focus on measurements and amounts. They tend to be based on numerical measurements of specific aspects of phenomena. They assume a fixed and measurable reality. The analysis is conducted by using numerical comparisons and statistical inferences and the data are reported through statistical analysis. The goal is to perform measurements and analyses that can be easily replicated by other researchers [54] [35].

Quantitative research procedure

The quantitative research approach starts with a specific theory, which is either proposed or previously developed, which leads to a specific hypothesis. One has to determine basic questions which should be answered by the study. Afterwards the participants have to be determined concerning the sample size and group. The next step is to do some research on existing literature as a literature review. After deciding which methods are needed to answer the questions the data are collected objectively and systematically by using for example surveys, correlation studies or field experiments. Afterwards data is measured quantitatively and evaluated using selected analysis tools. The final step is understanding and interpreting the results [52] [43].

Reasons for quantitative research design

- used for aspects of social behavior which can be quantified and patterned
- large sample size (can be even generalized to a whole population or sub-population)
- less time consuming in comparison to qualitative research
- reliability-research results can be reproduced [44]

3.3 Overview: Qualitative versus Quantitative Research

Table 3.1 displays the overview of the comparison between the qualitative and quantitative research approach.

	Qualitative	**Quantitative**
Conceptual	Concerned with understanding human behavior from the informant's perspective	Concerned with discovering facts about social phenomena
	Assumes a dynamic and negotiated reality	Assumes a fixed and measurable reality
Methodological	Data are collected through participant observation and interviews	Data are collected through measuring things
	Data are analyzed by themes from descriptions by informants	Data are analyzed through numerical comparisons and statistical inferences
	Data are reported in the language of the informant	Data are reported through statistical analyses

Table 3.1: Overview: Qualitative vs. Quantitative Approach [35]

3.4 Qualitative Research

Qualitative researchers study things in their natural setting and try to understand or interpret phenomena in terms of the meanings people give them. A dynamic and negotiated reality is assumed. Qualitative research involves using a variety of empirical materials like case studies, observations, personal experiences, interviews, or historical and document analyses. The analysis of the data is based on themes from the informants' descriptions, and the data are reported in the language of the informants [54] [35].

3.4.1 Qualitative research procedure

The first step when conducting a qualitative research procedure is to define some research questions. The spectrum is very diverse, but most of them focus on understanding meanings and social life in a particular context. As a starting point the researcher uses previous research, findings and theories and conducts a literature review. Another important point is to decide how the information is collected and how data analysis is chosen. The research data are collected using a specified approach (interview, observation,...) and notes and protocols are transcribed and prepared for analysis. After the analysis of the data, the final step is understanding and interpreting the results [36].

3.4.2 Reasons for qualitative research design

- used when the study has a specific contextual focus
- used when the study is about life experiences of a concept or phenomenon experienced by one or more individuals
- used to study areas in which there is little knowledge
- small sample size
- hard to reach target group
- tracking unique or unexpected events
- as a combination with quantitative approaches (as a preparation or completion) [36] [49]

3.5 Thinking aloud protocols

Thinking-aloud or Think-aloud protocols are a widely used qualitative method in usability testing. This method asks the participants to verbalize their thoughts while performing a task which is recorded on paper, audio, or video for further analysis. In many cases, the method of thinking aloud is a unique source of information about cognitive processes. There are two general types of thinking-aloud studies:

- Concurrent/Introspective Think Aloud
- Retrospective Think Aloud

The difference between those two methods is, that concurrent think aloud wants the user to comment his or her thoughts while solving the task and retrospective think aloud wants the user to complete one task and afterwards express his or her thoughts [22] [50].

The theoretical basis of Ericsson and Simon from 1993 [16] is most often cited, but in practice the work habits clearly diverge from this model. Sources for detailed descriptions on how to perform thinking aloud protocols show inconsistencies, while others do not even describe the thinking-aloud practices at all. Ericsson and Simon's approach is known to be very conservative. Their only focus is to consider what the participants attend to do and in what order. They explicitly exclude any data concerning the feelings of a participant as well as they want the participants to constantly speak aloud, otherwise they will be reminded to do so after 15-60 seconds of silence. The communication with the participants should be as short and nondirective as possible, and at best case, the participants should feel like "being alone in the room". The task flow of the participants should under any circumstances be disturbed, that is also why even neutral questions are not allowed. Boren et al. conducted field observations which showed that those reminders to think aloud often result in participants apologizing and feeling interrupted in their task flow, as well as this way of reminding makes the researcher seem controlling and authoritarian, while the participant feels inferior [5].

3.5.1 Relaxed thinking aloud protocols

There is a variant of thinking aloud, which constitutes a relaxation of the protocols proposed by Ericsson and Simon called "relaxed thinking aloud" or "interactive thinking aloud" which is most commonly used in practical usability testing. During relaxed thinking aloud the moderator prompts more often than in the classic thinking aloud, with the aim of motivating the participants to provide detailed verbalizations about their thoughts. Simon and Ericsson categorize verbalizations at 3 levels. Level 1 is simply the verbalization of information, which does not need to be transformed to report this information. Level 2 verbalizations are nonverbal like for example visual stimuli or movement and need to be transformed in verbal code. Level 3 verbalizations include the participants thoughts, ideas, hypotheses or motives and explanations. Level 3 also includes any outside influence like comments or prompts from the moderator. Simon and Ericsson just concentrate on level 1 and level 2 verbalization and ignore level 3 verbalization, while relaxed thinking aloud includes levels 1-3. This means, that reasoning, explanations, and the participants reflections are included. According to some sources, relaxed thinking aloud can influence the thought process and change user behavior. On the other hand, classic thinking aloud is said to be less informative due to the lack of explanations and reflections from the user [5] [60] [26].

Zhao et al. [63] compared the value of verbalization for detecting usability problems for classic and relaxed thinking aloud. The results suggest that relaxed thinking aloud leads to more valuable verbalizations for usability testing.

For our work, we are using the method of concurrent relaxed thinking aloud, as we expect this to give us the best output. This means the participants are thinking aloud during solving the tasks using a relaxation of the initially introduced conservative version of Ericsson and Simon.

3.6 Mixed Methods Research

Mixed-methods research is a research design or method used to collect, analyze, and combine quantitative and qualitative data in one study to better understand the research question. This approach has gained a lot of popularity in the last years because research methodology is constantly evolving, and mixed methods is another step forward that leverages the strengths of both qualitative and quantitative research [52]. Depending on the context, it is possible to gain more insights by combining qualitative and quantitative research than by either form alone. [12].

This approach is also called "Triangulation" and can for example combine the use of survey data with interviews. The main argument is that there should not be a contradiction between these two methodologies, but rather that it should be possible to bring them together [41].

There are different approaches of using mixed research methods. On the one hand there are sequential designs, which means the research is done in 2 phases were qualitative and quantitative data are collected in sequence.

Figure 3.1 shows the explanatory mixed methods design that involves a quantitative data collection phase followed by a qualitative data collection. This design is used to follow up on quantitative results by going more into depth through qualitative data.

Figure 3.1 also shows the exploratory mixed methods design in which qualitative data collection is followed up with quantitative data collection. The typical usage of this design is when quantitative instruments need to be developed when variables are unknown, or to examine preliminary qualitative results from a small group of individuals with a randomly selected sample from a larger population [52].

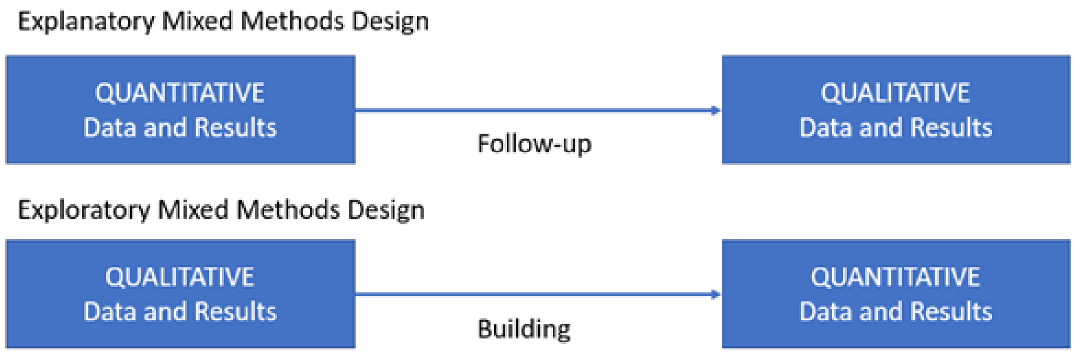

Figure 3.1: Two types of sequential designs [52]

On the other hand, one can conduct the mixed methods approach in parallel at the same time. In the triangulation mixed methods design, which is displayed in Figure 3.2 both quantitative and qualitative data are collected simultaneously, so one can converge the data to make comparisons and interpretations. This design is utilized for the comparison of the particular with the general or for validation of quantitative data with qualitative data.

The nested mixed methods design which is also displayed in Figure 3.2 is a slight variation of the triangulation design. Again, the quantitative and qualitative data are collected simultaneously,but with less emphasis on one of the two aspects. Also, the research questions vary and address different constructs. An example for this design would be when the researchers' goal is to conduct a randomized control experimental trial with the expected results to understand the impact of an intervention on outcomes, but also try to understand the process that participants go through during the study [52].

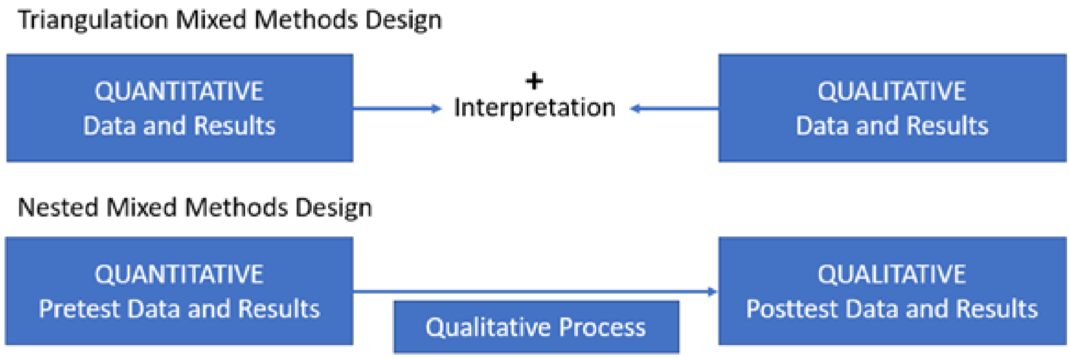

Figure 3.2: Two types of concurrent designs [52]

3.7 Reasoning

The reason why we decided to use both qualitative and quantitative methods is that for the specified research questions we wanted be able to gain as much information as possible and we saw the combination of both methods as an optimum.

CHAPTER 4

Experiment Setup

In this thesis we designed an online survey using “Limesurvey” [46] which displayed different pairs of directed acyclic graphs which were compared by the participants. The participants had to mark the perceived differences between the graph pairs by using a drag and drop functionality and placing markers on the chosen spots.

4.1 Colorblindness

Before participants had to complete the actual survey, they were subjected to a color blindness test. Color vision deficiency (CVD) affects approximately 1 in 12 men (8%) and 1 in 200 women (0,5%) in the world. People with a normal color vision, so called “trichromats”, use all three types of light cones in their eyes correctly, while the light cones in colorblind people eyes work faulty and perceive light differently. Depending on which light cone type does not work properly, there are three different types of colorblindness. The most common anomalous condition is “deuteranomaly”, a reduced sensitivity to green light followed by “protanomaly”, a reduced sensitivity to red light and the rare condition of “tritanomaly”, which is a reduced sensitivity to blue light [1].

We used “Ishihara color plates”[27], introduced by Dr.Shinobu Ishihara for testing, which display numbers, which should be clearly visible for viewers with a normal color vision, while people with an underlying color vision deficiency may see different numbers or nothing. In figure 4.1 you can see an example of a Ishihara plate. People with a normal color vision can see the number "8", while people with a red-green deficiency see the number "3".

Depending on the output of our colorblindness test, the participants were redirected to either the “normal vision survey” or the “colorblind-friendly survey”.

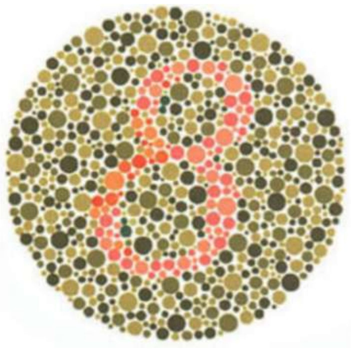

Figure 4.1: Ishihara test plate; normal vision:8, red-green deficiency:3 [27]

4.2 Color Selection

We used “Colorbrewer 2.0” [11] to create a colorblind safe selection of colors for the graphs. Colorbrewer recommended 3 colors which are displayed in figure 4.2.

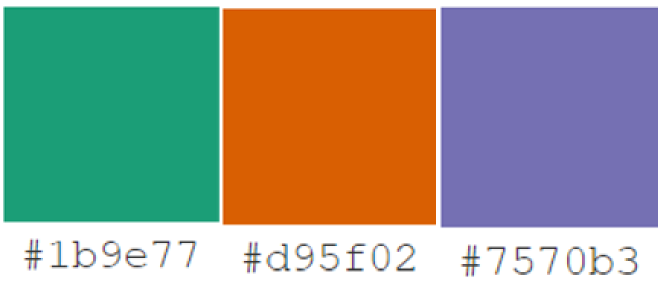

Figure 4.2: Colobrewer recommendation for 3 colorblind friendly colors displayed in colorhex.com [10]

We additionally tested the chosen colors using a view filter in the image editor “Gimp” [19], which simulates different color vision deficiencies. Figure 4.3 displays how the colors are perceived by people with normal color vision, protanopia, deuteranopia or tritanopia.

For the participants with normal color vision, we used the color selection which can be seen on Figure 4.4.

The reason for choosing a different color scheme for people with a normal vision was because of some feedback of test subjects, who perceived the chosen colors as not that distinguishable. We conducted a small survey, which included 10 participants with normal color vision, which had to choose between the 2 different color palettes. The results showed that 9 participants voted for the color selection with the “yellow-like” color, while 1 participant voted for the color palette including the “orange-like”color.

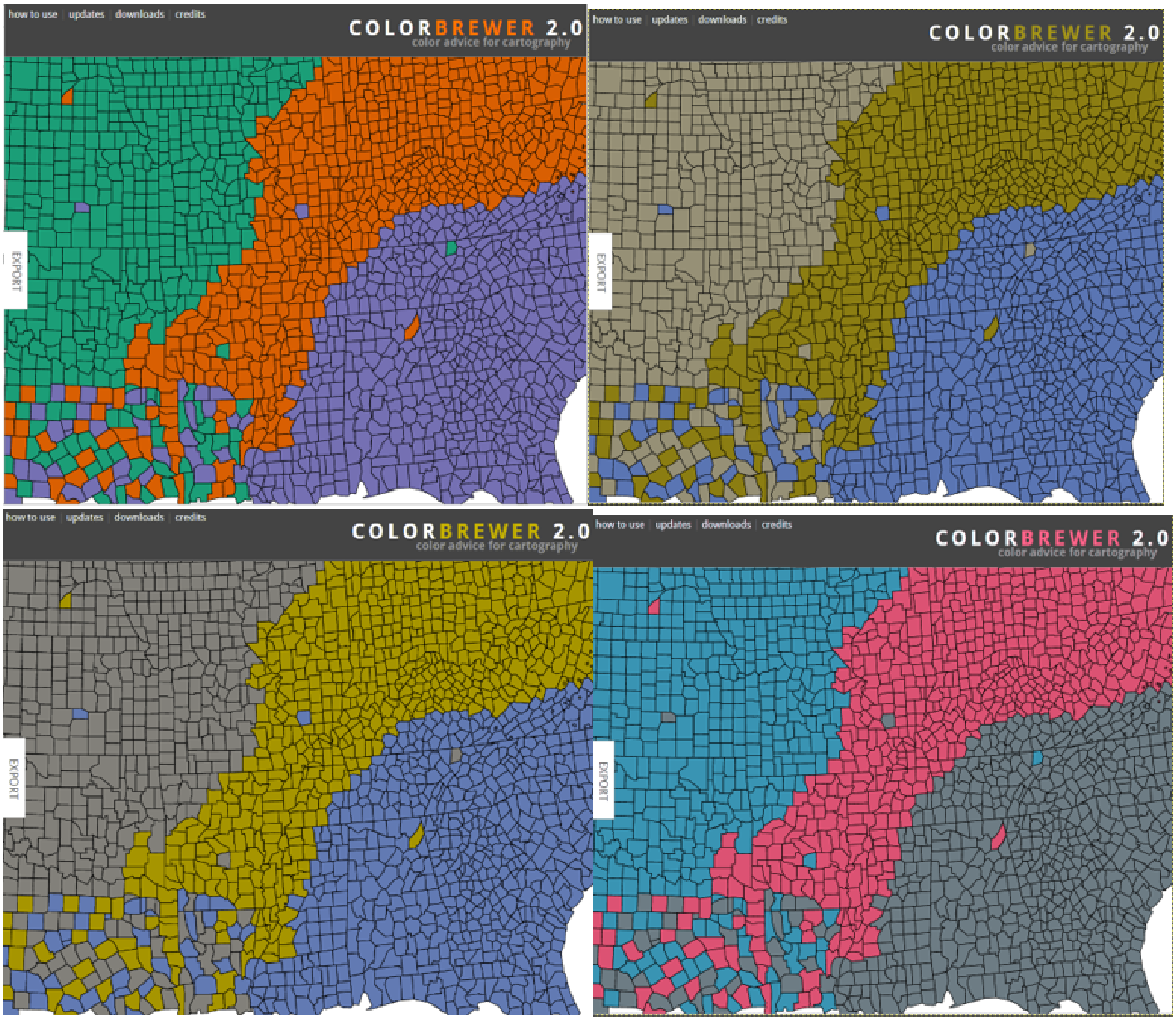

Figure 4.3: top left: normal CV, top right: protanopia, bottom left: deuteranopia, bottom right: tritanopia [11]

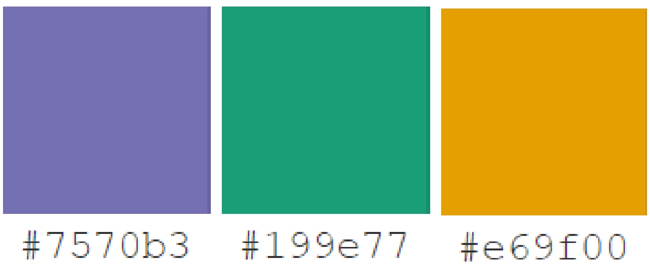

Figure 4.4: color selection for normal vision [10]

When choosing colors for a visualization, which should display different data classes, it is important that the chosen colors are differentiable. The international commission on illumination (CIE) was the first to address the topic of color difference in 1976 and established a standard color distance metric called “Delta E”. In the following years, the color difference equation got adapted and improved by considering more attributes, which may influence the perception. Delta E 76 was followed by Delta E94 and finally Delta E 2000, which is actually the new industry standard [15].

To confirm our color selection, we calculated Delta E 2000 for our colors using an online calculator [9].

The higher the Delta E value, the more differentiable are the colors. A value of Delta E $<= 1$ means a difference not perceptible by human eyes, while the value 100 shows that the colors are exactly the opposite [59].

As we can see in table 4.1, our chosen colors are approved being differentiable enough to be used for our experiment.

Colors	Delta E
	53,928568
	44,419326
	39,383637
	42,481106
	54,725156

Table 4.1: Delta E calculation for chosen colors

On figure 4.5 one can see the comparison between the graph with the normal sight color selection and a graph with the colorblind-friendly color selection.

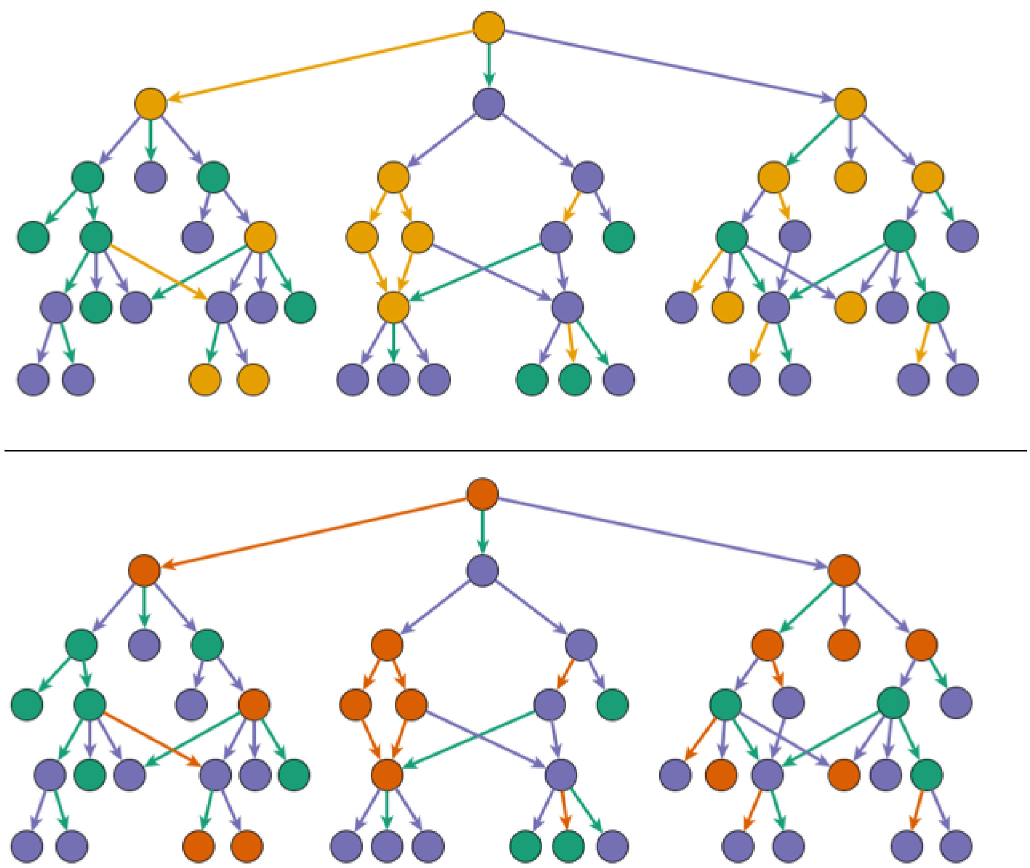

Figure 4.5: **top:** normal sight graph version; **bottom:** colorblind-friendly graph version

4.3 Hypotheses

Based on previous research, we defined the following hypotheses:

Whitespace

Hypothesis 1: Differences are more likely to be found where white space is.

Edge length

Hypothesis 2: Differences are more likely to be found in changes to longer edges.

Multicolored or uniform surrounding

Hypothesis 3: Differences are more likely to be found on the background of uniform edges and nodes.

Position: Outside/Inside

Hypothesis 4: Differences are more likely to be found on the outside.

Size of the graph (small or large)

Hypothesis 5: Differences are more likely to be found in small graphs.

4.4 Experiment Description

The survey started with a declaration of consent, which had to be accepted by the participants, followed by basic demographic questions regarding the age, gender and familiarity with visualizations measured on a 5-point scale (1…very familiar, 5… very unfamiliar). The survey consisted of 16 graph pairs including 8 smaller graphs (40-56 nodes and 44-61 edges) and 8 larger graphs (97-104 nodes and 104-124 edges) and 2 test graphs at the beginning. For comparability with previous studies, we used a very similar dataset as Wallner et al[58]. Each survey question displayed a base graph and the altered version above. The differences in the altered graph did influence the colors of nodes and/or links and did not change the graph structure. Participants had to mark the perceived differences by using drag and drop to place a marker on the desired spot.

The graphs were displayed in a semi-random order. We had 6 different sequences to avoid similar changes being displayed consecutively.

At the beginning of the survey there were 2 test questions, which were designed to familiarize oneself with the system by testing the drag and drop functionality and eventually adjusting the zoom factor of the screen, so that one could see the graphs on full screen. The first test question had no time limit whereas the second one had a 3-minute limit.

The experiment was split into two phases and two groups, the quantitative part, and the qualitative part.

Group one (Phase 1) consisted of 41 participants who had to solve the survey under time constraint. For each graph pair the participants had either 40 or 60 seconds, depending on the graph size (larger or smaller graph).

In the quantitative part, after every graph pair, the participants were asked how certain they were about finding all differences and how difficult it was to find the differences. The possible answers were measured on a 5-point scale reaching from 1. . . very certain; 1. . . very easy to 5. . . very uncertain; 5. . . very difficult.

The second group (Phase 2) included 10 participants who had to solve the same survey but without a time limit imposed, as the focus was on the participants' explanations using a relaxed thinking aloud method.

In the qualitative part, the participants just had to answer how difficult they perceived each graph pair, which was measured like in the quantitative part with a 5-point scale.

Several pilot tests were conducted with computer science students from the University of Cologne, which helped to determine the level of difficulty and correct timing, and to obtain feedback on the overall study design.

4.5 Participant Selection

Besides for example, hypotheses, methods, procedures, and methods of analysis also the participant selection plays an important role when it comes to the reproducibility of research results. For this visualization evaluation we selected computer science students as participants. Studies concerning the application of information visualizations in the medical domain compared the results of physicians with those of students and showed, that the differences were not highly significant. Previous work shows that students are very good candidates to identify usability issues as they take more time to complete the tasks. And as the human cognition, like for example color vision, visual search or perception is very similar for all humans, students are also suitable, e.g., for evaluating general cognitive processes. Furthermore, it is easier to recruit students than experts [29].

4.6 Graph creation

For comparability we designed graphs similar to the ones used in previous studies conducted by Guenter Wallner et al [58] [57]. For the creation we used yEd graph editor [61]. The 16 graphs could be split in 4 different categories of complexity and in 2 groups of graph sizes. The complexity of the graphs was derived from the number of edges and nodes as well as from the size of the graphs. Table 4.2 displays the graphs with their complexity level and size.

Graph	Complexity (1-4)	Size (big/small)
B1G1	1	small
B1G2	1	small
B1G3	1	small
B1G4	1	small
B2G1	2	small
B2G2	2	small
B2G3	2	small
B2G4	2	small
B3G1	3	big
G3G2	3	big
B3G3	3	big
B3G4	3	big
B4G1	4	big
B4G2	4	big
B4G3	4	big
B4G4	4	big

Table 4.2: Graph complexity and size

The 16 graph pairs included between 2 and 5 differences each which can be seen in table 4.4. We had 1 graph containing 2 differences, 7 graphs containing 3 differences, 6 graphs containing 4 differences and 2 graphs containing 5 differences, which is visualized in table 4.3.

Number of differences	Number of graphs
2	1
3	7
4	6
5	2

Table 4.3: Number of graphs with same amount of differences

Graph	Number of differences
B1G1	2
B1G2	3
B1G3	3
B1G4	3
B2G1	3
B2G2	3
B2G3	3
B2G4	3
B3G1	4
G3G2	4
B3G3	4
B3G4	4
B4G1	4
B4G2	4
B4G3	5
B4G4	5

Table 4.4: Number of differences per graph

4.7 Factors

For every difference we defined 5 factors: graphsize (big/small), surrounding color (uniform/multicolored), whitespace (yes/no), position (inside/outside), edge length (no edge/short/long). The above mentioned factors were varied systematically among the graphs.

4.7.1 Graphsize

The factor graphsize could take either the value “small” or “big”. 8 graphs were considered to be “small” and had between 40-56 nodes and 44-61 edges, while the other 8 graphs were declared to be “big” and included between 97-104 nodes and 104-124 edges which is displayed in table 4.5.

Graph	Number of nodes	Number of edges	Size
B1G1	52	59	Small
B1G2	52	59	Small
B1G3	56	61	Small
B1G4	56	61	Small
B2G1	40	45	Small
B2G2	40	45	Small
B2G3	43	45	Small
B2G4	42	44	Small
B3G1	97	105	Big
G3G2	97	105	Big
B3G3	99	104	Big
B3G4	100	105	Big
B4G1	100	119	Big
B4G2	100	119	Big
B4G3	104	121	Big
B4G4	104	124	Big

Table 4.5: Number of nodes and edges per graph

4.7.2 Color

The factor color describes if the surrounding edges and nodes of the difference were uniformly colored or multicolored. In figure 4.6 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. This difference would have the property: color=uniform.

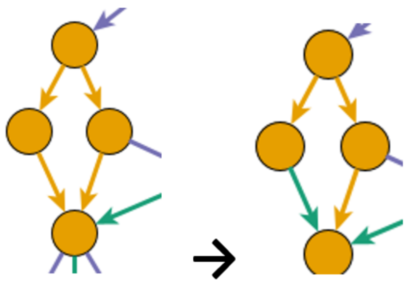

Figure 4.6: Example of an uniform surrounding

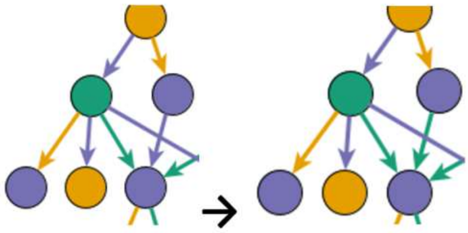

Figure 4.7: Example of an multicolored surrounding

In figure 4.7 one can see the that the violett edge changes in a green edge, but the surrounding is not uniform, therefore this difference would be categorized as: color=multicolored.

4.7.3 Whitespace

As for the factor whitespace we defined that there must be space in minimum 2 directions. We had very specific guidelines to define whether a difference has whitespace or not. Differences at the very left and very right side where not automatically defined as having much whitespace, as the images of the graphs had a certain size and in the study, they were not displayed on a white background, but a colored one. To be defined as having whitespace our measure was that there had to be space, 2 times as big as the difference, in 2 of the 4 positions (top, right, bottom, left). In figure 4.8 you can see an example. Difference A has space at the top and at the left side (2/4 positions), difference B has eventually space just on the left (1/4) and difference C has no space in any kind of direction (0/4). As a result, difference A was defined having whitespace, while difference B and C were defined as having no whitespace.

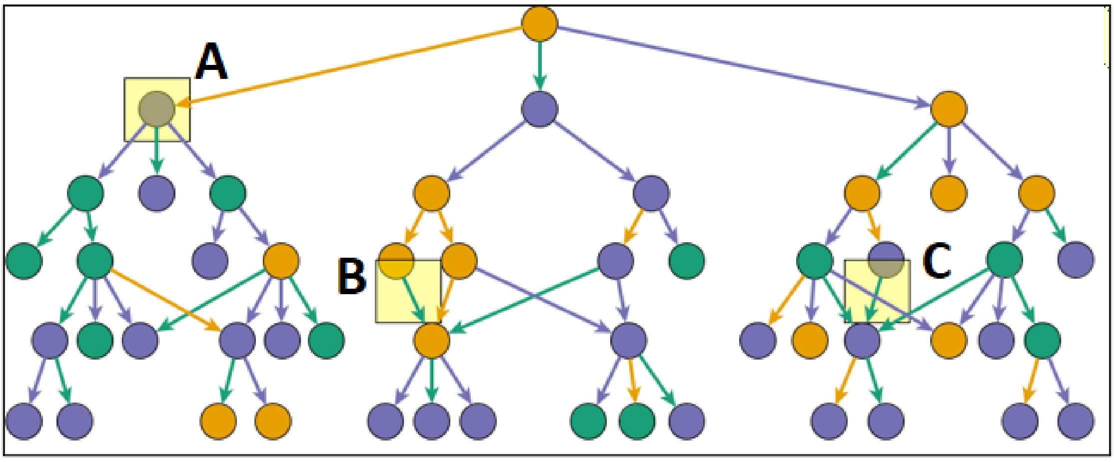

Figure 4.8: Example of defining whitespace

4.7.4 Position

To define whether a difference is placed on the inside or on the outside of a graph, we specified, that every difference which is not at the hull of the graph, is considered to be a difference on the inside. On figure 4.9, the differences A and D would be defined as outside, while the differences B and C would be at the inside of the graph.

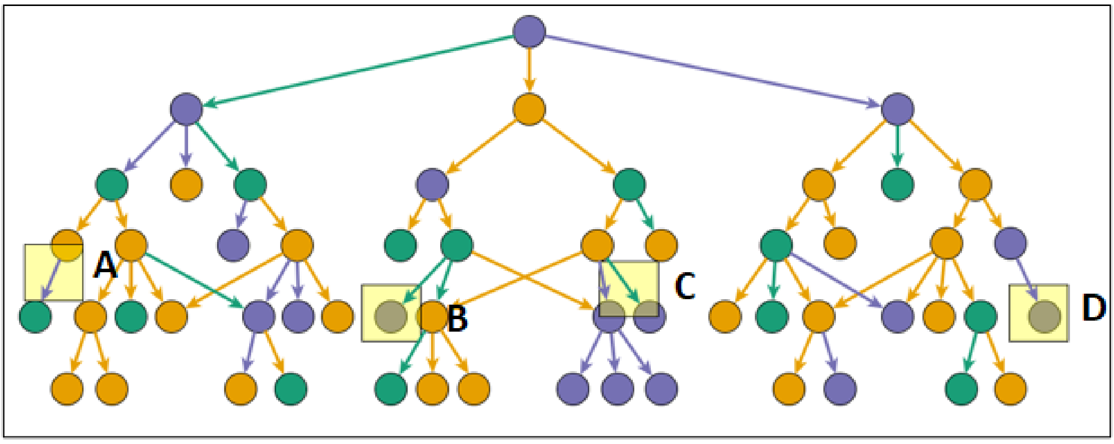

Figure 4.9: Example of defining position

4.7.5 Edge length

For defining if an edge was long or short, we did not have one specific measure but specified it for each graph individually. We took a look at the longest edges and at the shortest to decide about the categories. The difference between a short and a long edge was set very differentiable, to avoid confusion. If the difference was a node, there was obviously no edge length, and the factor was 0. Figure 4.10 shows a short edge, which was difference A and in comparison, a long edge at difference C.

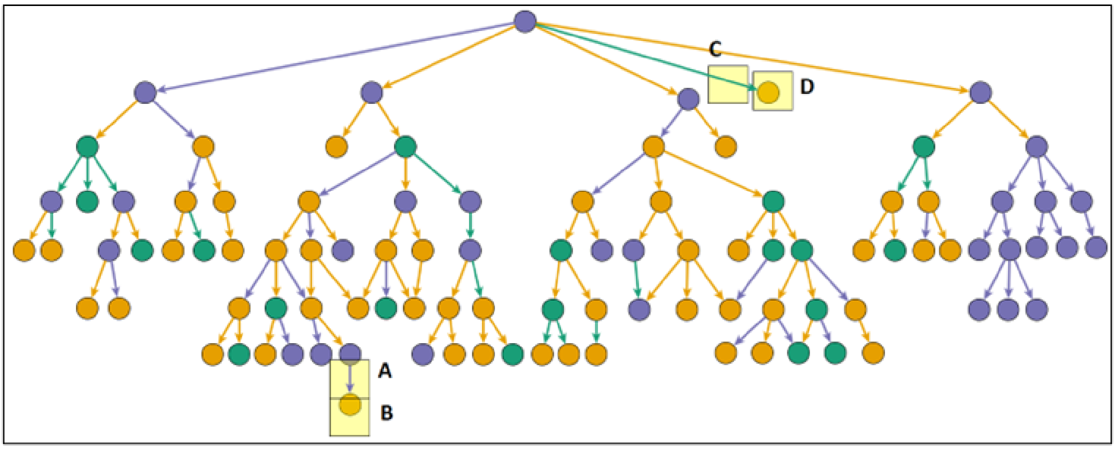

Figure 4.10: Example of defining edge length

CHAPTER 5

Results

5.1 Results Quantitative Part

5.1.1 Participants

In total we received 41 complete responses from 11 females and 30 males. The average age of the participant was 25 years, whereas the youngest was: 22 years and the oldest 29 years.

The participants rated their familiarity with visualizations on average with 2.5, based on a 5-point scale (1...very familiar, 5...very unfamiliar). The average time to complete the survey were 22 minutes.

5.1.2 Difficulty and Certainty

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 possible answers were measured on a 5-point scale reaching from 1... very easy to 5... very difficult and 1... very certain to 5... very uncertain. In table 5.1 you can see the average difficulty and certainty displayed for every graph sorted in descending order. Graph B4G1 was on average rated to be not only the most difficult graph but participants rated on average to have the highest uncertainty of finding all the differences. On the other hand, graph B2G2 was considered to be the easiest graph and also had the highest certainty of finding all the differences.

Graph	Avg Difficulty		Graph	Avg Certainty
B4G1	4		B4G1	3,87804878
B3G1	3,78048781		B4G3	3,87804878
B4G3	3,78048781		B3G1	3,75609756
B4G2	3,65853659		B4G2	3,65853659
B4G4	3,41463415		B4G4	3,31707317
G3G2	3,34146342		G3G2	3,2195122
B3G3	3,02439024		B1G1	3
B1G1	3		B1G2	2,82926829
B3G4	3		B3G4	2,82926829
B1G2	2,73170732		B2G3	2,80487805
B1G3	2,48780488		B1G3	2,73170732
B1G4	2,48780488		B1G4	2,70731707
B2G1	2,41463415		B3G3	2,70731707
B2G3	2,41463415		B2G1	2,29268293
B2G4	2,31707317		B2G4	2,07317073
B2G2	2,07317073		B2G2	1,97560976

Table 5.1: Average difficulty and certainty per graph (quantitative results)

Spearman Correlation

In R we calculated some Spearman correlations for our data. The Spearman correlation coefficient value ranges from -1 to +1, while -1 is a perfectly negative association between two variables, 0 indicates no association between the values and the value of +1 indicates a perfectly positive association. In other words, a negative correlation means, that if one variable increases, the other one decreases and a positive correlation means that if one value increases, the other one also tends to increase. In the Spearman Correlation calculation, we also get a p-value, which indicates if the results are statistically significant or not. A p value less than 0.05 indicates significance [18].

Correlation between found differences and difficulty

First, we calculated the correlation between found differences and perceived difficulty.

```
>
> cor.test(data$percFound, data$percDiff, method = "spearman", exact=FALSE)

        Spearman's rank correlation rho

data:  data$percFound and data$percDiff
S = 964.97, p-value = 0.1062
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.4190689
```

Figure 5.1: Spearman Correlation between found differences and difficulty

As we can see in illustration 5.1, the p value is higher than 0.05, which indicates that the results are not significant.

Correlation between found differences and certainty

We calculated the Spearman correlation between found differences and perceived certainty.

```
>
> cor.test(data$percFound, data$percCert, method = "spearman", exact=FALSE)

        Spearman's rank correlation rho

data:  data$percFound and data$percCert
S = 1056.2, p-value = 0.02624
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.5531758
```

Figure 5.2: Spearman Correlation between found differences and certainty

As we can see in image 5.2, the p-value is lower than 0.05 which indicates a significance. The Rho value, also called the spearman correlation coefficient is -0.553, which indicates a negative association between found differences and perceived certainty. Which means, if the value of found differences increases, the uncertainty decreases.

Correlation between certainty and difficulty

We calculated the Spearman correlation between certainty and difficulty.

```
>
> cor.test(data$percCert, data$percDiff, method = "spearman", exact=FALSE)

        Spearman's rank correlation rho

data:  data$percCert and data$percDiff
S = 50.259, p-value = 2.653e-07
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9260904
```

Figure 5.3: Spearman Correlation between certainty and difficulty

The results which can be seen in illustration 5.3 indicate, that the results are significant and that there is a positive correlation between certainty and difficulty. If the perceived uncertainty increases also the perceived difficulty increases.

We categorized the 16 graphs in 4 categories depending on their basis structure. B1 included B1G1, B1G2, B1G3 and B1G4 and so on. We calculated the average certainty and difficulty for each of the four groups. For repetition, the answers were measured on a 5-point scale reaching from 1. . . very easy to 5. . . very difficult and 1. . . very certain to 5. . . very uncertain.

Looking at the results in the table 5.2 and diagram 5.4, there might be a dependance between graph complexity and perceived difficulty and certainty. The results do not display a linear growth, but one can notice that starting with complexity group 2 until 4, the perceived average difficulty increases and the perceived uncertainty of having found all the differences increases.

	G1	**G2**	**G3**	**G4**
(Un)Certainty	2,817	2,287	3,128	3,683
St.dev.(un)certainty	0,133	0,37	0,472	0,265
Difficulty	2,677	2,305	3,287	3,713
St.dev.difficulty	0,244	0,161	0,364	0,244

Table 5.2: Average difficulty and certainty per complexity group and the standard deviation

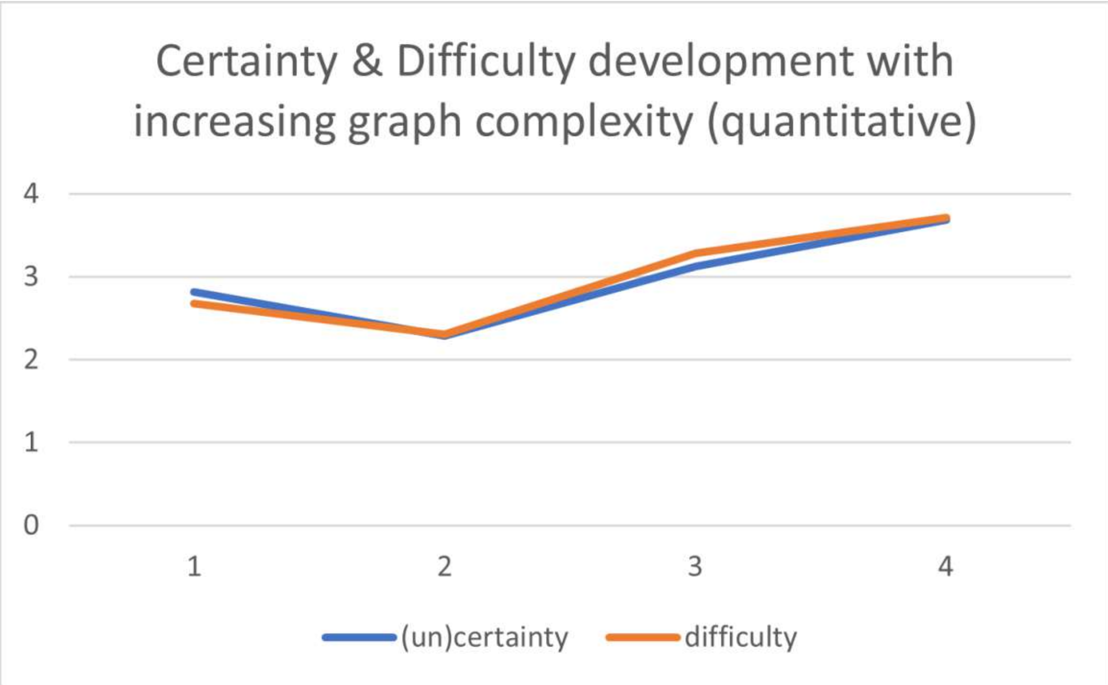

Figure 5.4: Certainty and difficulty development with increasing graph complexity

5.1.3 Evaluation

In “Limesurvey” the participants had to place markers by using drag and drop to mark the perceived differences. For this functionality we used a custom question theme which we adapted for our purpose and included in Limesurvey, which was running on the university server ambrose.iguw.tuwien.ac.at. Limesurvey saved the placed coordinates relative to the image size. For evaluation we wrote a script in python and executed it in “Jupyter Notebook”, a web-based interactive computing platform. Before running the script, we had to prepare the data, as for example, the coordinates were saved in reverse order and the exported data had additional, unused columns. The code used the prepared and filtered excel data coordinates from the participants and placed a dot for each coordinate on the respective graph image. For each of the 16 graphs + 2 test graphs we executed the code and jupyter created an image with the marked differences for each participant for each graph. Figure 5.5 displays the python code.

```
excelData = pd.read_excel('Results.xlsx',index_col=0)
c1 = excelData["TQ2/Coordinates1"]
c2 = excelData["TQ2/Coordinates2"]

#range 0 to number of participants
for i in range(0, 41):
    #print(str(c1.tolist()[i]) + " - " + str(c2.tolist()[i]) + " - " + str(c3.tolist()[i]))
    #print(x)
    plt.figure(figsize=(10, 10))
    plt.rc('font', size=15)
    name = 'TQ2-' + str(i+1)
    plt.text(10,50, name)

    #depending on how many differences there are in a graph, there are as many coordinateSets
    coordinateSet1 = str(c1.tolist()[i]).split(",")
    if coordinateSet1[0] != "nan" and coordinateSet1[0]!= "undefined":
        xCoordinate1 = float(coordinateSet1[0])
        yCoordinate1 = float(coordinateSet1[1])
        plt.plot(10*xCoordinate1*x,10*yCoordinate1*y, marker='o', color="black", markersize=15)
        plt.text(10*xCoordinate1*x,10*yCoordinate1*y,'  Q1')


    coordinateSet2 = str(c2.tolist()[i]).split(",")
    if coordinateSet2[0] != "nan"and coordinateSet2[0]!= "undefined":
        xCoordinate2 = float(coordinateSet2[0])
        yCoordinate2 = float(coordinateSet2[1])
        plt.plot(10*xCoordinate2*x,10*yCoordinate2*y, marker='o', color="black", markersize=15)
        plt.text(10*xCoordinate2*x,10*yCoordinate2*y,'  Q2')


    plt.imshow(data)
    Save to file
    plt.savefig(name+'.png')
```

Figure 5.5: Displays python code for generating images with the placed coordinates

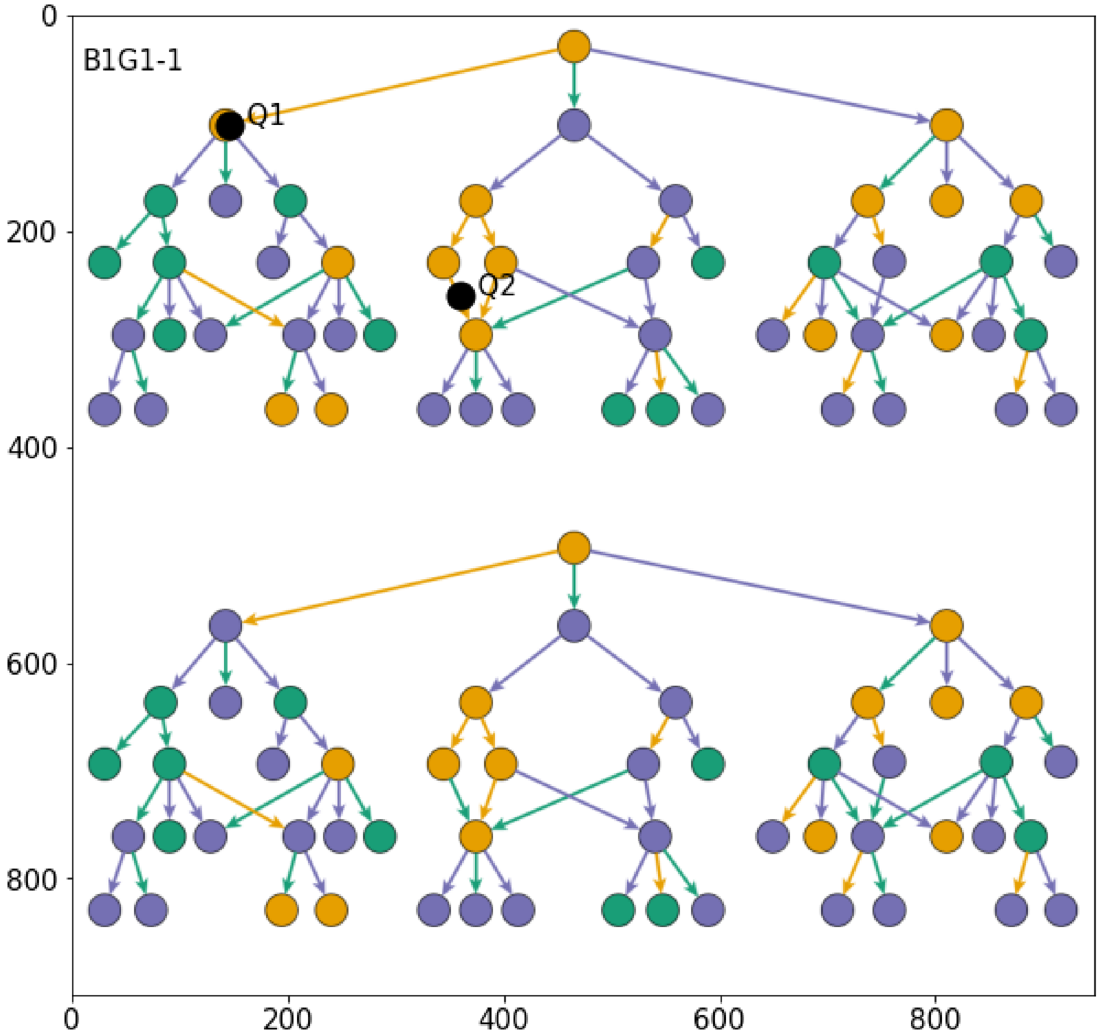

Figure 5.6: Image created by the python code displaying the marked differences in graph B1G1 for participant 1 including the order in which they were perceived, e.g. Q1= first marked difference.

On figure 5.6 you can see an example output of our python code. The final outcome was a collection of all graphs for every participants. marked with his or her found differences.

After we generated the images, we had to design a solution template for each graph. We named each difference individually using letters, like one can see in figure 5.7. At the attachment you can find the whole table displaying every single difference and its factors.

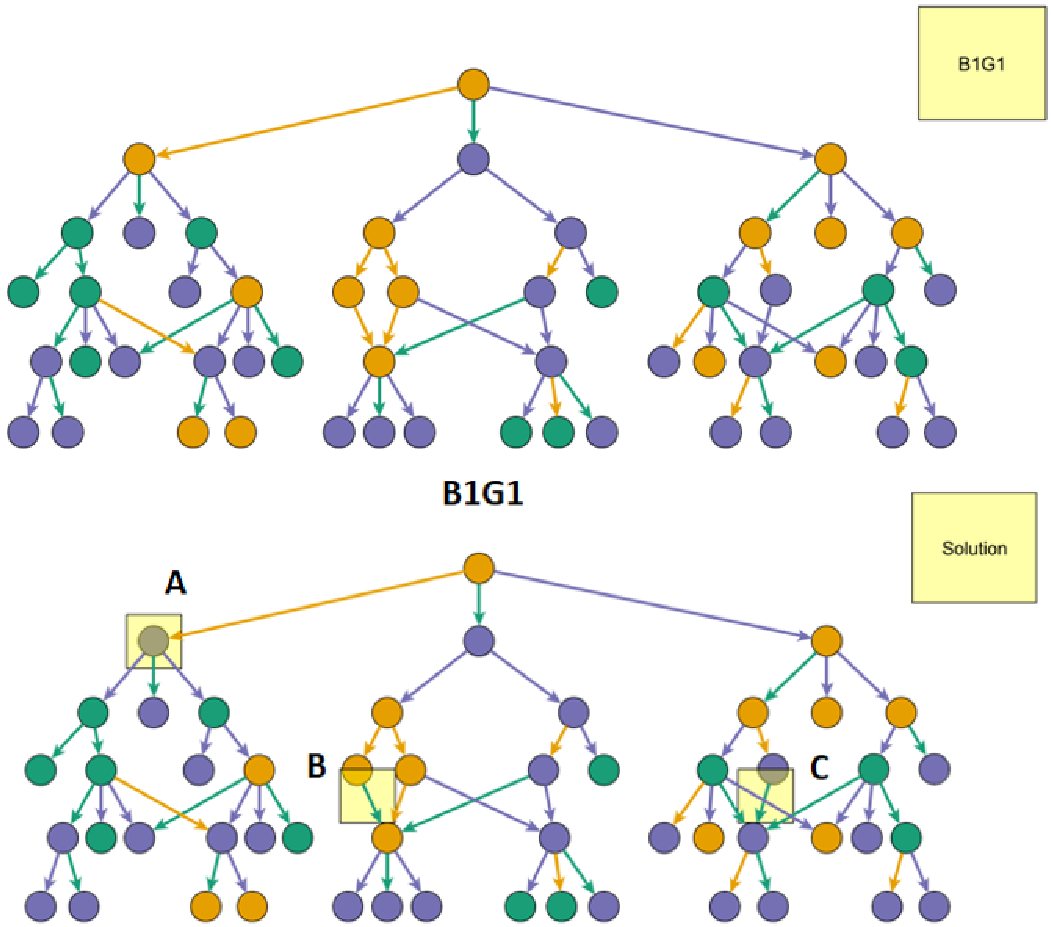

Figure 5.7: Solution template for graph B1G1

Having this template designed for all graphs, we manually compared the jupyter notebook output images with the template images and could identify if a difference was found or not. Furthermore, we were able to infer the order in which the differences were perceived.

5.1.4 Chi-square Test

Having our data prepared, we conducted a chi-square test of independence, which is a statistical hypothesis test to determine if two categorical or nominal variables are likely to be related or not [8]. For every calculation one defines a null hypothesis. For the result to be significant, the p-value must be less than 0.05. If the p-value is less than 0.05, one can reject the null hypothesis. Our proposed null hypotheses:

$H_0 g$: Graph size and found differences are independent.
$H_0 p$: Position and found differences are independent.
$H_0 w$: Whitespace and found differences are independent.
$H_0 c$: Color and found differences are independent.
$H_0 e$: Edge length and found differences are independent.

Factor	p	X2	Cramer‘s V
Graph size	0,001972	9,58	0,064009409
Position	0,1036	X	X
Whitespace	2,8368E-05	17,524	0,086594324
Colored/uniform	0,00016552	14,187	0,07791367
Edge length	0,01031426	6,58	0,145804929

Table 5.3: Results of the chi square test for quantitative data

The results of our calculation, which can be seen in table 5.3 show that there is a relationship between the graph size and found differences, between whitespace and found differences, between color and found differences and between edge length and found differences, while there is no relationship between the position and found differences. We can reject all null hypotheses besides $H_0 p$. The Cramer's value indicates how strong the perceived effects are and ranges from 0 to 1. In our case the effects are quite weak.

Hypothesis	Statement	Result
Hypothesis 1	Differences are more likely to be found where white space is.	accept
Hypothesis 2	Differences are more likely to be found in changes to longer edges.	accept
Hypothesis 3	Differences are more likely to be found on the background of uniform edges and nodes.	accept
Hypothesis 4	Differences are more likely to be found on the outside.	reject
Hypothesis 5	Differences are more likely to be found in small graphs	accept

Table 5.4: Hypotheses status after quantitative evaluation

Looking at our proposed hypotheses, we can say, that all hypotheses besides hypothesis 4, which is about the position of the differences, are true, which is visualized in table 5.4 .

We also conducted a generalized estimated equation (GEE) in R for the 5 factors. When

```
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 5.8: Results GEE in R

looking at the results in figure 5.8, we first take a look at the Pr(<|W|) value, which tells us how well each predictor variable is able to predict the value of the response variable in the model. If the value is less than 0.05, the predictor variable is seen as significant. In our case, position is the only value which is not less than 0.05, but rather 0.97088, which means position is not significant. This again proves our before mentioned statements.

To see the influence of the predictors, we have to look at the first column with “Estimates”. For example, an one unit increase in the predictor variable size, is associated with an average change of -0.23655 in the log odds of the response variable “found” taking on a value of 1. This means that a higher value of size (and in our case size=1 means a big graph) is associated with a lower likelihood of the variable “found” taking on the value of one (found=1, means difference found). So in other word- a bigger graph decreases the probability of finding a difference [62].

5.2 Sequences

Next, we analyzed the sequences in which the differences were perceived. As our sequences vary, as there are between 2 and 5 differences per graph, we have decided to take a closer look at the first and last found differences. We assumed that differences found first by the participants would be particularly noticeable and immediately catch the eye and provide information about the importance of each factor, which later on could be used for facilitating graph drawings.

Concerning the quantitative data, we analyzed which differences were perceived most commonly as first difference and which most commonly as last for every graph. For each graph we then had the most common first/last difference, and we listed all the according factors of these differences. Afterwards we grouped the differences in our earlier described factor-categories and calculated for each category the total sum of the first or last most common perceived differences. Afterwards we could deduce that the most common first seen difference was one of the type with long edges, followed by a uniform surrounding as second most common first difference and differences positioned inside, while the most common last difference were with whitespace, followed by short edges, and differences positioned on the outside of the graph. Table 5.5 and table 5.6 display the ranking of the most common first perceived differences and the most common last perceived differences. As a disclaimer we have to add, that looking at the last perceived difference in the quantitative data, might not be as significant, as there was a time limit. We therefore cannot say if it was influenced by lack of time or was the hardest to find. We still did want to take these results into account.

Ranking	**Type of difference**	**# of differences per type**	**For 41 participants**	**Found as first difference**	**Percent**
1.	long edge	8	328	98	29,88
2.	uniform surrounding	16	656	185	28,20
3.	Position: inside	27	1107	174	15,72

Table 5.5: Ranking of the most common first difference (quantitative)

Ranking	**Type of difference**	**# of differences per type**	**For 41 participants**	**Found as last difference**	**Percent**
1.	Whitespace: yes	21	861	106	12,31
2.	Position: outside	30	1230	151	12,28
3.	Short edge	31	1271	142	11,17

Table 5.6: Ranking of the most common last difference (quantitative)

To additionally approve our percentual calculations, we conducted a binary logistic regression in SPSS. We adapted our dataset and added the value “first”, which could be 0 or one, depending on if the perceived difference was found first or not.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	size(1)	,021	,102	,040	1	,841	1,021	,836	1,246
	position(1)	-,112	,117	,916	1	,338	,894	,710	1,125
	whitespace(1)	-,096	,126	,577	1	,447	,908	,709	1,164
	color(1)	,764	,105	52,874	1	<,001	2,146	1,747	2,637
	edgelength			56,622	2	<,001			
	edgelength(1)	-,200	,117	2,902	1	,088	,819	,651	1,031
	edgelength(2)	,856	,156	29,982	1	<,001	2,354	1,733	3,199
	Constant	-1,256	,143	77,556	1	<,001	,285		

a. Variable(s) entered on step 1: size, position, whitespace, color, edgelength.

Figure 5.9: SPSS output for binary logistic regression for first seen difference

Illustration 5.9 shows in the “Sig.” column, which factors are significant and which not. Every p-value below 0.05 is seen as significant. Furthermore the column “Exp(B)” shows us the Odds Ratio, which makes a statement about the extent to which the presence or absence of one characteristic is related to the presence or absence of another characteristic and how strong this relationship is.

- Odds > 1 indicate that it is more likely that the event will occur than that it will not occur.
- Odds of 1 indicate that the occurrence and non-occurrence of the event are equally likely.
- Odds < 1 indicate that it is more likely that the event will not occur than that it will occur [25].

In our example the significant values are color =1 (which means for our data “uniform surrounding”) with the p-value: <0.001 and the Exp(B) value 2.145 and edge length=2 (which means for our data “long edges”) with the p-value: <0.001 and the Exp(B) value 2.345. 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. Those values are comparable with our previous percentage calculations and confirm the statement.

5.3 Results Qualitative Part

5.3.1 Participants

In total we conducted thinking aloud protocols with 8 participants including 2 females and 6 males. The average age of the participants was 25 years, whereas the youngest was: 18 years and the oldest 26 years. The participants rated their familiarity with visualizations on average with 3.5, based on a 5-point scale (1... very familiar, 5... very unfamiliar). The average time to complete the survey was not taken under account, as the qualitative part did not include a time constraint.

5.3.2 Difficulty

In the qualitative part, after each graph pair, the participants were asked how difficult it was to find the differences. The possible answers were measured on a 5-point scale reaching from 1... very easy to 5... very difficult. In comparison to the quantitative part, we left out the question about how certain the participants were that they found all the differences. In table 5.7 you can see the average difficulty perceived by the participants, displayed for every graph. Graph B4G1 was rated on average to be the most difficult one, while graph B2G2 was perceived as being the easiest one. The results are the same as for the quantitative part, where as well B4G1 was perceived as the most difficult and B2G2 was perceived as the easiest graph.

Graph	**Average difficulty**
B4G1	3,875
B4G2	3,75
B4G3	3,75
B4G4	3,625
G3G2	3,5
B3G3	3,125
B3G1	2,75
B3G4	2,75
B2G4	2,625
B1G2	2,375
B1G3	2,375
B1G1	2,25
B1G4	2
B2G3	2
B2G1	1,875
B2G2	1,5

Table 5.7: Displays the average difficulty per graph (qualitative results)

Spearman Correlation

Correlation between found differences and difficulty
We calculated the Spearman correlation for the percentage of found differences and the perceived difficulty.

```
> cor.test(data$percFoundQl, data$percDiffQl, method = "spearman", exact=FALSE)

        Spearman's rank correlation rho

data:  data$percFoundQl and data$percDiffQl
S = 1103.9, p-value = 0.009871
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6234263
```

Figure 5.10: Spearman Correlation between found differences and difficulty

As illustration 5.10 shows, the results are significant and the rho value of -0,623 indicates a negative correlation between the variables, which means that as the percentage of found differences increases, the difficulty decreases.

Analogous to the procedure for the quantitative data, we categorized the 16 graphs in 4 categories depending on their complexity and basis structure. Afterwards we calculated the average difficulty for each of the four groups. For example, to calculate G1 we took the average of the average difficulty averages from B1G1, B1G2, B1G3 and B1G4, and continued with the procedure for G2,G3 and G4, which is displayed in table 5.8.

Looking at the results in table 5.8 and diagram 5.11, there might be a dependance between graph complexity and perceived difficulty. The results do not display a linear growth, but one can notice that starting with complexity group 2 until 4, the perceived average difficulty increases.

	G1	**G2**	**G3**	**G4**
Difficulty	2,25	2	3,031	3,75
St.dev.difficulty	0,177	0,468	0,359	0,102

Table 5.8: Average difficulty per complexity group and the standard deviation

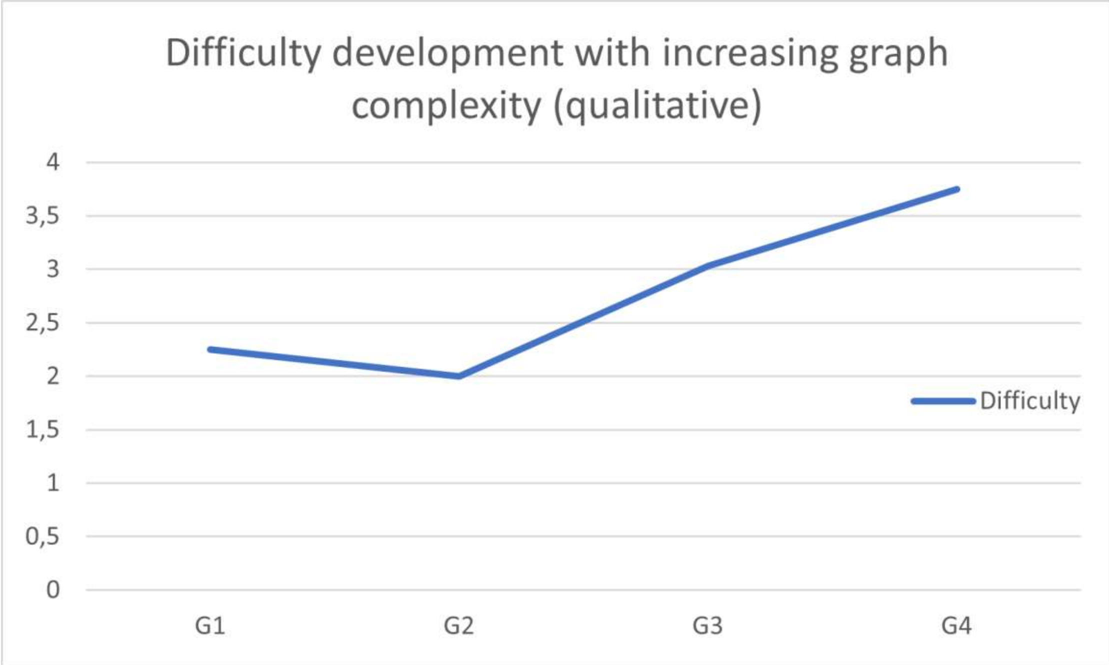

Figure 5.11: Difficulty development with increasing graph complexity (qualitative)

We designed a template for the thinking aloud procedure of our participants in phase 2, the qualitative part. In table 5.9 we can see the individual steps, which were a guideline for conducting the thinking aloud protocols.

Step	Description
1.	Introduction, explanation of the topic,problem statement, usage
2.	Declaration of concept has to be accepted
3.	Explanation of the procedure
4.	Colorblindness test
5.	Start of the normal or colorblind friendly survey
6.	2 test questions to get familiar with the drag and drop functionality; explanation: survey should be done without longer breaks; no time constraint, but time is not unlimited for every graph pair
7.	Participant is getting informed to please think aloud while solving the tasks Thinking aloud should contain: how one is proceeding, where did one start, what is easy/difficult, any other comments
8.	The participant decides when he or she is ready to move on, or he or she spends more than 3,5 minutes on that comparison- then he or she will be informed to move on
9.	After each graph pair the participant has to answer how difficult he or she found the task on a 5 point scale
10.	Afterwards he or she gets asked "Was there anything especially difficult/interesting for you" "Do you want to add any feedback/comment?" and gets reminded to think aloud while proceeding
11.	Steps 7.-10. get repeated for all 16 graphs.
12.	At the end of the survey the participant can leave feedback or a comment and can describe his or her overall impressions.

Table 5.9: Template for the qualitative evaluation procedure

After evaluating the thinking aloud data, we could identify 3 different strategies, that were used to compare the graph pairs. Every participant used a combination of strategies and did not stick to only one strategy. We tried to classify the strategies in 3 categories:

- Hierarchical
- Layer-by-layer
- Separating nodes and edges

The hierarchical approach included splitting the graph in sub-trees, going from left to right, from right to left, using a top to bottom approach, and using a bottom to top approach. The "layer-by-layer" strategy was to horizontally look at the nodes and edges and compare them going layer by layer, as the name indicates. Participants using the "separating nodes and edges" strategy, first concentrated just on finding differences looking at the nodes and afterwards on the edges, or the other way round. As already mentioned above, every participant used more than one strategy. Table 5.10 shows the different strategies and how many participants used them.

Strategy	Sub-Strategy	Number of participants that used this method
Hierarchical	Split into subtrees	5 (P1; P2; P3; P6; P7)
	left→right	7 (P1; P1; P3; P5; P6; P7; P8)
	right→left	2 (P4; P5)
	top→ bottom	8 (P1; P2; P3; P4; P5; P6; P7; P8)
	bottom→ top	4 (P1; P6; P7; P8)
Layer-by-Layer		3 (P6; P7; P8)
Separating nodes & edges		5 (P1; P3; P4; P5; P6)

Table 5.10: Different strategies

5.3.3 Chi-square test

For the qualitative data we also calculated a chi-square test which can be seen in table 5.11. Analogous to the quantitative version, the p-value provided information whether a factor was significant or not. For the qualitative part the p-value was below 0.5 for graph size and color, which means that there is a relationship between the graph size and the found differences as well as a relationship between the colored/uniform surrounding and found differences. Position, whitespace and edge length were not significant.

Factor	p	X2	Cramer‘s V
Graph size	0,000783191	11,28	0,157
Position	0,347804622	X	X
Whitespace	0,128622864	X	X
Colored/uniform	0,036932818	4,35	0,098
Edge length	0,066299243	X	X

Table 5.11: Results chi square test qualitative data

Looking at our proposed hypotheses in 5.12, we can say, that just the hypothesis concerning the graph size and colored/uniform surrounding is supported.

Hypothesis	Statement	Result
Hypothesis 1	Differences are more likely to be found where white space is.	reject
Hypothesis 2	Differences are more likely to be found in changes to longer edges.	reject
Hypothesis 3	Differences are more likely to be found on the background of uniform edges and nodes.	accept
Hypothesis 4	Differences are more likely to be found on the outside.	reject
Hypothesis 5	Differences are more likely to be found Differences are more likely to be found in small graphs	accept

Table 5.12: Hypotheses status after qualitative evaluation

5.3.4 Sequences

For the qualitative data, we also analyzed which differences were perceived first and which last for every graph. Analogous to our procedure with the quantitative data, we also listed all the factors of each most common perceived differences and could observe that the most common first seen difference were long edges and differences with a uniform surrounding, while the most common last difference were differences with no whitespace and differences positioned in the middle of the graph.
Table 5.13 and 5.14 display the ranking of the most common first perceived differences and the most common as last perceived differences for the qualitative data.

Ranking	**Type of difference**	**# of differences per type**	**For 8 participants**	**Found as first difference**	**Percent**
1.	long edge	8	64	30	46,86
2.	uniform surrounding	16	128	34	26,56
3.	Whitespace: no	36	288	68	23,61

Table 5.13: Ranking of the most common first difference (qualitative)

Ranking	Type of difference	# of differences per type	For 8 participants	Found as last	Percent
1.	Whitespace: no	36	288	64	22,22
2.	Position: inside	27	216	47	21.76
3.	multicolored surrounding	41	328	67	20,43

Table 5.14: Ranking of the most common last difference (qualitative)

During the thinking aloud process, we were also able to filter out what was particularly difficult and what was particularly easy when looking for differences.

What was perceived as being difficult was edge crossing, because the participants claimed, that edge crossings disturb them while trying to follow the paths of the graphs. Nearly every participant said that his or her concentration level lowered with time and that it got harder to look for differences. When there was no or little whitespace it was also harder for the participants to distinguish differences, as well as when 2 differences were next to each other, because one did not expect that. Furthermore, participants claimed, that the bigger the graph got, the more confusing it became and that finding faulty edges was way more difficult than finding nodes. In addition, the participants declared that they found it hard to distinguish between the violet color and the green color.

On the other hand, participants perceived that if there was whitespace surrounding the difference, they could more easily detect it, as well as when the difference was at the outside of the graph, so for example at the very top or bottom or at the sides. Furthermore many participants added, that it was easier to detect nodes in comparison to edges, the only exception was if the faulty edge was long, then it was also easy to detect it. The participants also mentioned that they perceived it easy if a difference was in a uniform surrounding, as they could easily and quickly detect it and could move on with checking the graph more quickly, because they could skip the uniformly colored parts. After some time and after finishing some graph comparisons, the interviewed participants were more proficient, according to their own statements.

5.3.5 Thematic analysis

To further analyze the data, we conducted a thematic analysis. This method is best suitable for identifying and analyzing patterns of meaning in a data set. The output of the data should display the most salient patterns of the content [23].
The first step is to examine the whole data, to be able to develop a coding frame and to subsequently code the data. We developed our coding frame which contained: Statements concerning the strategy, statements concerning the difficulty, statements concerning what was perceived as easy and other statements, which can be seen in table 5.15.

Code name	**Definition**	**Example**
Strategy	Statement concerning the procedure of looking for differences	"Going from top to bottom"
Difficulties	Verbalizations about occurrences which were perceived as being difficult or which made the process of searching more complicated	“Edge crossing made it difficult to follow the branches”
Easy	Verbalizations about occurrences which were perceived as being easy or which facilitated looking for differences	“Whitespace helped with spotting the differences”
Others	Other verbalizations concerning the overall study	“If no errors found, compulsively searching for errors”

Table 5.15: Coding frame for the verbalizations extracted from the thinking aloud protocols

Even if participants mentioned their statements a couple of times, it was just counted as one. The table shows which verbalization was used at least once by whom of the 8 participants. We counted the verbalizations falling with the different codes, which is displayed in table 5.16 and could indicate their importance.

Strategy	P1	P2	P3	P4	P5	P6	P7	P8	C
From left to right	X	X	X		X	X	X	X	7
From right to left				X	X				2
Top to bottom	X	X	X	X	X	X	X	X	8
Bottom to top	X					X	X	X	4
View nodes and edges separately	X		X	X	X	X			5
Split in subtrees	X	X	X			X	X		5
Search layer by layer, horizontally		X	X	X		X	X	X	6
Search for symmetrical differences								X	1
Mouse for orientation	X	X							2
Remember colors and quantity	X	X					X	X	4
Recite color	X						X		2
When checking: look over chaotically				X		X	X		3
When checking: move from bottom to top	X							X	2
Difficulties	P1	P2	P3	P4	P5	P6	P7	P8	C
No whitespace	X								1
Edges harder to detect	X				X				2
The more elements, the more difficult	X		X	X		X		X	5
2 differences next to each other	X			X					2
Green and purple harder to distinguish	X		X						2
Edge crossing	X	X	X	X	X		X	X	7
Concentration decreases with time			X	X	X		X		4
Easy	P1	P2	P3	P4	P5	P6	P7	P8	C
Difference: very bottom/top/ “sticking out”	X	X						X	3
Nodes easier to detect	X					X			2
Uniform surrounding	X		X	X	X		X	X	6
More practiced after some time	X		X	X	X		X		5
Whitespace between elements					X			X	2
Long edges					X	X		X	3
Other	P1	P2	P3	P4	P5	P6	P7	P8	C
"Compulsively" search for errors if no found	X			X			X	X	4
Larger graph= more likely to check again	X								1
If graph easier= less motivation to check				X			X		2

Table 5.16: Categorized verbalizations

Strategy

Concerning the strategy, all 8 participants used the top to bottom approach at least once during the study, while 7 out of 8 participants were looking for differences starting from left to right. 6 out of 8 participants searched layer by layer going through the graphs horizontally. In the table 5.17 you can see the individual verbalizations regarding the strategy together with the percentage of how many of the 8 participants mentioned it at least once, sorted in descending order.

Strategy	**Percent**
Top to bottom	100
From left to right	87,5
Search layer by layer, horizontally	75
View nodes and edges separately	62,5
Split in subtrees	62,5
Bottom to top	50
Remember colors and quantity	50
When checking: look over chaotically	37,5
From right to left	25
Mouse for orientation	25
Recite color	25
When checking: move from bottom to top	25
Search for symmetrical differences	12,5

Table 5.17: Strategy verbalizations in descending order

Difficulties

Most of the participants, 7 out of 8, perceived edge crossing as being difficult when looking for differences. The bigger the graph, the harder it gets, was claimed by 5 out of 8 participants. 50 percent of the participants mentioned that their concentration decreased over time. Further difficulties are displayed in the table 5.18.

Perceived as difficult	**Percent**
Edge crossing	87,5
The more elements, the more difficult	62,5
Concentration decreases with time	50
Edges harder to detect	25
2 differences next to each other	25
Green and purple harder to distinguish	25
No whitespace	12,5

Table 5.18: Difficulty verbalizations in descending order

Perceived as easy

A uniform surrounding was perceived as easy by 6 out of 8 participants. 5 out of 8 participants claimed, that after some time they were more practiced compared to the beginning. 37,5 percent of the participants stated that long edges helped with finding differences. In table 5.19 you can see the verbalizations categorized as easy and the according percentage.

Perceived as easy	**Percent**
Uniform surrounding	75
More practiced after some time	62,5
Long edges	37,5
Difference: very bottom/top/ “sticking out”	37,5
Nodes easier to detect	25
Whitespace between elements	25

Table 5.19: Verbalizations of what was perceived as easy in descending order

Other verbalizations

Half of the participants said that they started to compulsively search for errors if they could not find any in an area. 25 percent of the participants added, that if they perceived a graph as being easy, their motivation to check the graph again was lower. On the other hand, 2 out of 8 participants claimed, that they were more likely to check the graph again when it was a larger graph. The percentages can be taken from the table 5.20.

Other	**Percent**
"Compulsively" search for errors if none found	50
If graph easier= less motivation to check	25
Larger graph= more likely to check again	12,5

Table 5.20: Other verbalizations in descending order

5.4 Summary Results

After our research work, we are now able to answer our research questions, which we defined in advance.

5.4.1 Research Question 1

R1) Do the variable color and the factors whitespace, position, uniform/multicolored background, graph size and edge length influence the recognition of differences in DAGs (under time constraint)?

Answer:
Quantitative Data:
The factors whitespace, uniform/multicolored background, graph size and edge length influence the recognition of differences in DAGs while position has no influence on the recognition under time constraint.

Qualitative Data:
The factors uniform/multicolored surrounding and graph size influence the recognition of differences in DAGs, while position, whitespace and edge length have no influence on the recognition without any time constraint.

5.4.2 Research Question 2

R2)
Do the variable color and the factors whitespace, position, uniform/multicolored background, graph size and edge length affect the sequence of perceived differences of DAGs? Are some variables perceived earlier than others?

Answer:
Quantitative Data:
The most common first seen difference was one of the type with long edges, followed by a uniform surrounding as second most common first difference, while the most common last difference was placed where there is whitespace, followed by differences positioned on the outside of the graph.

Qualitative Data:
The most common first seen difference in the qualitative part, were long edges and differences with a uniform surrounding, while the most common last difference were differences with no whitespace and differences positioned in the middle of the graph.

5.4.3 Research Question 3

R3) Which strategies are used to compare two graphs?
Answer:
We identified 3 different categories including:

- Hierarchical
- Layer-by-layer
- Separating nodes and edges

5.5 Discussion

Quantitative:
Our quantitative study revealed that the factors whitespace, edge length, uniform/ multicolored surrounding, and graph size are significant, while the position is not significant. No statistically significant interactions effect between the factors could be observed. According to our accepted hypotheses we can say that differences are more likely to be found where there is whitespace, in changes with longer edges, changes with a uniform surrounding and in small graphs. On the other hand, we can say that our hypothesis of differences being more likely to detect on the outside of a graph got rejected.
Wallner et al. [57] could observe similar results about whitespace. In their study whitespace was also mentioned to be helpful when perceiving differences. In the work from Wallner et al. [58], they observed the factor shape, especially outer shape, and the results indicated that the importance of the outer shape was not as important compared to the other factors. In our case the outer shape could be compared with our factor “position”, where we define if a difference is placed inside a graph or on the outside. Our results showed that position has no significant influence on the perception.

The chi squared test indicated, that graph size is a significant factor when perceiving differences, which could be compared with the results from Ballweg et al. [4], where the number of levels (depth) and the number of nodes on a specific level dominantly influenced similarity perception.

Most common as first difference perceived factors were long edges and a uniform surrounding. The uniform surrounding could be compared with Wallner et al. [57] study were the participants formed groups of nodes, which they could easily detect. In our work the uniform surrounding made the differences more present in the graph and such changes were in consequence often perceived at first.
The most common as last perceived differences were differences with whitespace and differences positioned on the outside. We must add, however, that looking at the most common last perceived differences in the quantitative part is not as meaningful because time constraint might play a role. If a graph was perceived as being more difficult, the uncertainty of having found all differences also increased. There might be a dependance

between graph complexity and perceived difficulty as the results do not display a linear growth, but one can notice that starting with complexity group 2 until 4, the perceived average difficulty increases. Similar results could be observed in Wallner et al. [58], were with increasing graph complexity, also the uncertainty and difficulty increased.

The Spearman correlations between certainty and difficulty and found differences and certainty were significant, while the correlation between found differences and difficulty was not significant. Wallner et al. [58] could calculate significant correlations between all three combinations.

Qualitative:
Our study has revealed 3 different search strategies including hierarchical, layer-by-layer and separating nodes and edges. Our results are comparable with the ones from Wallner et al [57], who could distinguish the layer-by-layer, hierarchical and mixture of both strategies. Like in the previous study, the top-to-bottom approach was used most commonly. While Wallner et al. [57] , could observe an alternated top to bottom approach, where participants went down a branch and came back up the next branch, we could not observe anything similar in our study.
Similar to the previous work, we could just observe 2 out of 8 participants going from right to left, whereas most of the participants preferred to search from left to right. Every participant used more than one strategy and varied the chosen ones over the graph pairs. Like the results from Wallner et al. [57] we can say that the choice of strategy depended on the overall graph structure and varied from graph to graph and participant.

A chi square test indicated that the surrounding and graph size are significant factors. The work from Wallner et al. [57] mentioned groups of nodes as factors, which helped the perception of differences. In our case this can be somehow compared with a uniform surrounding of nodes and edges, which is significant. The most common first found difference was one of the type: long edges, followed by differences in an uniform surrounding, while the most common as last perceived differences were differences with no whitespace and differences positioned inside the graph. There might be a dependance between graph complexity and perceived difficulty. The results do not display a linear growth, but one can notice that starting with complexity group 2 until 4, the perceived average difficulty increases.
The amount of found differences, or in general existing differences, had an influence on the certainty of having found all the differences. If participants just found, for example two differences, even though there were just two differences in the graph, they kept on checking the graph again more likely than when they already have found 4 differences. The cluelessness of how many differences there were per graph, led to uncertainty.
Edge crossing seemed to interfere with participants' search flow, as nearly every participant perceived edge crossing as difficult. Edge crossing made it difficult for participants to split the graph into subgraph. Similar behavior could be observed in previous work from Wallner et al. [57], were many participants stated that edge crossing complicated the procedure of searching. In contrary they also had participants mentioning that edge

crossing was useful, which was in our opinion, mainly to the fact that in their study edges and nodes appeared or disappeared while they stayed the same in our work and just the colors changed. The appearance of a newly added edge crossing might have been the reason for participants to judge edge crossing as helpful, which was also confirmed in a later work of Wallner et al. [58], were a newly introduced edge crossing helped to spot a change. Interestingly in the work from Ballweg et al [4] edges and edge crossing had no influence on the similarity perception and seemed not to matter to the participants.

Whitespace was perceived as being helpful when searching for differences. Additionally, participants claimed that differences "sticking out", like ones placed at the very top or bottom or on the sides, were more likely to perceived earlier than others. Furthermore, the uniform surrounding was perceived as helpful, because participants could go through a certain area of the graph more quickly, and differences appeared more dominant. While symmetry was the most common mentioned factor in the study of Wallner et al- [57], surprisingly in our study just one participant focused on searching for symmetric differences.

Half of our qualitative study participants stated at least once that they tried remembering colors and the quantity of edges and nodes and tried to recognize some patterns. Similar observations could be made in the work from Wallner et al. [57], were they defined this factor as "shape", and participants looked for recognizable, most often geometric, forms in the graphs.

While the majority of the participants stated that by the time, they got more familiar and trained with looking for differences, they also claimed that their concentration decreased with time.

In the quantitative part as well as in the qualitative part the graph B4G1 was rated to be the most difficult, and the graph B2G2 as the easiest one. Even though we used a semi random order (we had 6 different sequences which were randomly assigned to the participants), we cannot completely rule out the possibility that the graph pair order might have an influence.

5.6 Summary

To summarize, we could observe some similar results as Wallner et al. [57], Wallner et al. [58] and Ballweg et al. [4], but we could determine that the perception of differences is affected by more different factors, as well as we could distinguish more strategies.
The results showed that the factors whitespace, uniform/multicolored background, graph size and edge length influence the recognition of differences in DAGs while position has no influence on the recognition under time constraint. The most common first seen differences in the were of the type: long edges and uniform surrounding, for the quantitative part as well as for the qualitative part. Our found strategies included: hierarchical, layer-by-layer and separating nodes and edges. Our results provided new insights and results in visual

graph comparison, as well as some contradictions to previous work and can be used as a starting point for future work. Further research might concentrate on the certain colors used and their influence, as well as deeper work concerning the sequences in which differences are perceived as well as on the relationship between used strategies and type of difference.

APPENDIX A

Factors displayed for each graph

B1G1

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B1G1	A	small	outside	yes	multicolored	X
B1G1	B	small	inside	no	uniform	short
B1G1	C	small	inside	no	multicolored	short

B1G2

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B1G2	A	small	outside	no	uniform	short
B1G2	B	small	outside	no	uniform	X
B1G2	C	small	outside	no	multicolored	short

B1G3

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B1G3	A	small	outside	no	multicolored	short
B1G3	B	small	inside	no	multicolored	X
B1G3	C	small	inside	no	uniform	short
B1G3	D	small	outside	no	multicolored	X

B1G4

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B1G4	A	small	outside	yes	multicolored	X
B1G4	B	small	inside	yes	uniform	X
B1G4	C	small	outside	yes	multicolored	short
B1G4	D	small	inside	no	uniform	X

B2G1

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B2G1	A	small	inside	no	uniform	long
B2G1	B	small	outside	yes	multicolored	long
B2G1	C	small	inside	no	multicolored	short
B2G1	D	small	inside	no	multicolored	short

B2G2

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B2G2	A	small	inside	no	multicolored	long
B2G2	B	small	outside	no	multicolored	long
B2G2	C	small	inside	no	multicolored	long

B2G3

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B2G3	A	small	outside	yes	multicolored	short
B2G3	B	small	outside	yes	multicolored	X
B2G3	C	small	outside	yes	multicolored	short
B2G3	D	small	outside	yes	multicolored	short
B2G3	E	small	outside	yes	multicolored	X

B2G4

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B2G3	A	small	outside	yes	multicolored	short
B2G4	A	small	inside	no	multicolored	long
B2G4	B	small	outside	yes	multicolored	short
B2G4	C	small	outside	yes	multicolored	X

B3G1

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B3G1	A	big	inside	no	uniform	short
B3G1	B	big	inside	no	multicolored	short
B3G1	C	big	inside	no	uniform	short

B3G2

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B3G2	A	big	inside	no	uniform	short
B3G2	B	big	inside	no	uniform	short
B3G2	C	big	inside	no	multicolored	short

B3G3

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B3G3	A	big	outside	yes	multicolored	short
B3G3	B	big	outside	yes	multicolored	X
B3G3	C	big	inside	no	uniform	long
B3G3	D	big	inside	yes	multicolored	X

B3G4

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B3G3	A	big	outside	yes	multicolored	short
B3G4	A	big	outside	yes	multicolored	short
B3G4	B	big	outside	yes	multicolored	short
B3G4	C	big	outside	yes	uniform	short
B3G4	D	big	outside	yes	uniform	X

B4G1

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B3G3	A	big	outside	yes	multicolored	short
B4G1	A	big	inside	no	multicolored	short
B4G1	B	big	inside	no	multicolored	short

B4G2

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B4G2	A	big	outside	no	multicolored	short
B4G2	B	big	outside	no	multicolored	X
B4G2	C	big	outside	no	multicolored	short

B4G3

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B4G2	A	big	outside	no	multicolored	short
B4G3	A	big	outside	no	uniform	short
B4G3	B	big	inside	no	multicolored	X
B4G3	C	big	inside	no	multicolored	short
B4G3	D	big	outside	no	multicolored	short

B4G4

Graph	Coord	Size	Position	Whitespace	Surrounding	Short/long edge
B4G2	A	big	outside	no	multicolored	short
B4G3	A	big	outside	no	uniform	short
B4G4	A	big	outside	yes	uniform	X
B4G4	B	big	outside	yes	multicolored	X
B4G4	C	big	inside	no	multicolored	long
B4G4	D	big	inside	no	multicolored	short
B4G4	E	big	inside	no	multicolored	X

APPENDIX B

Graph pairs with solution

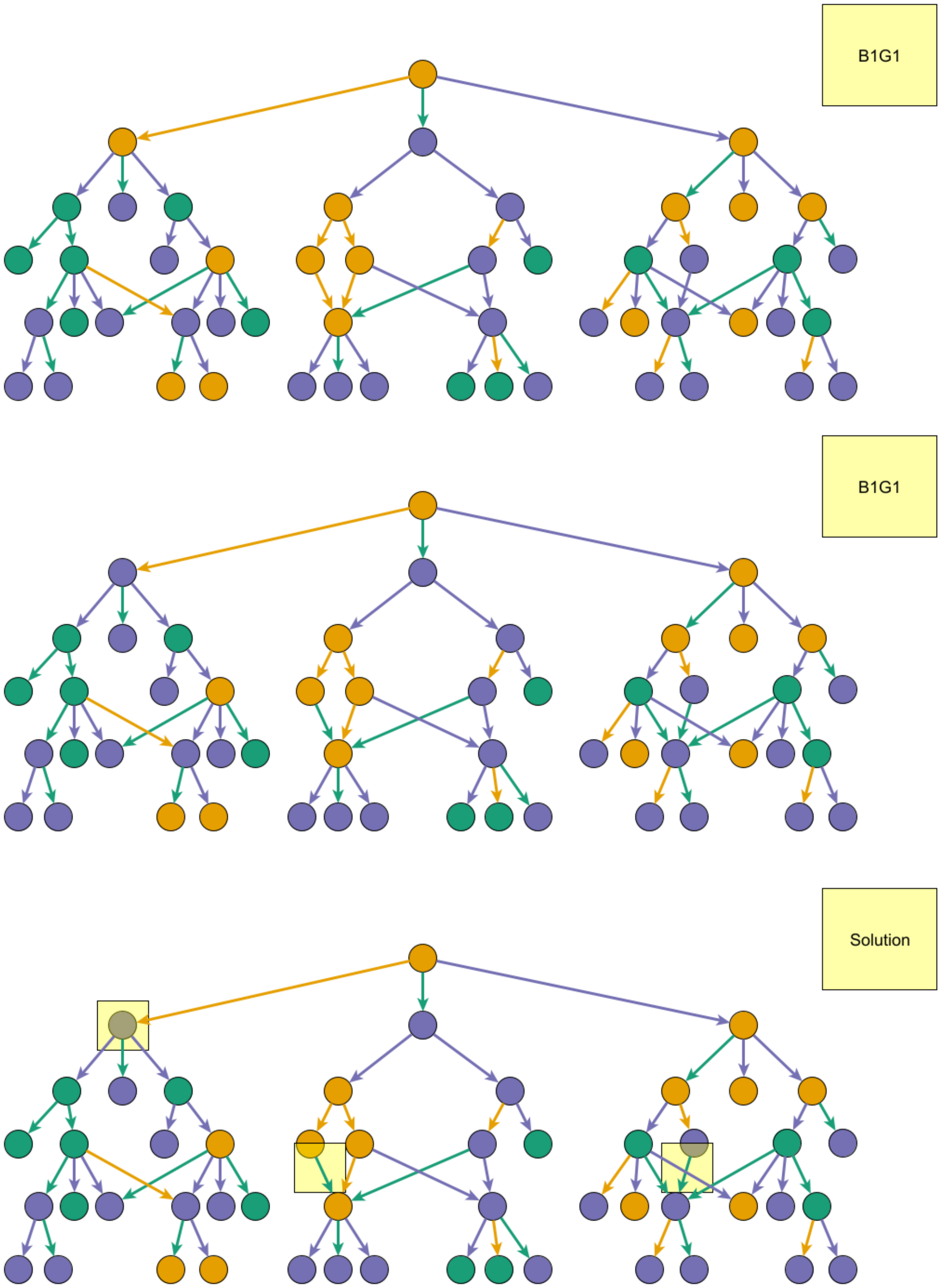

Figure B.1: B1G1

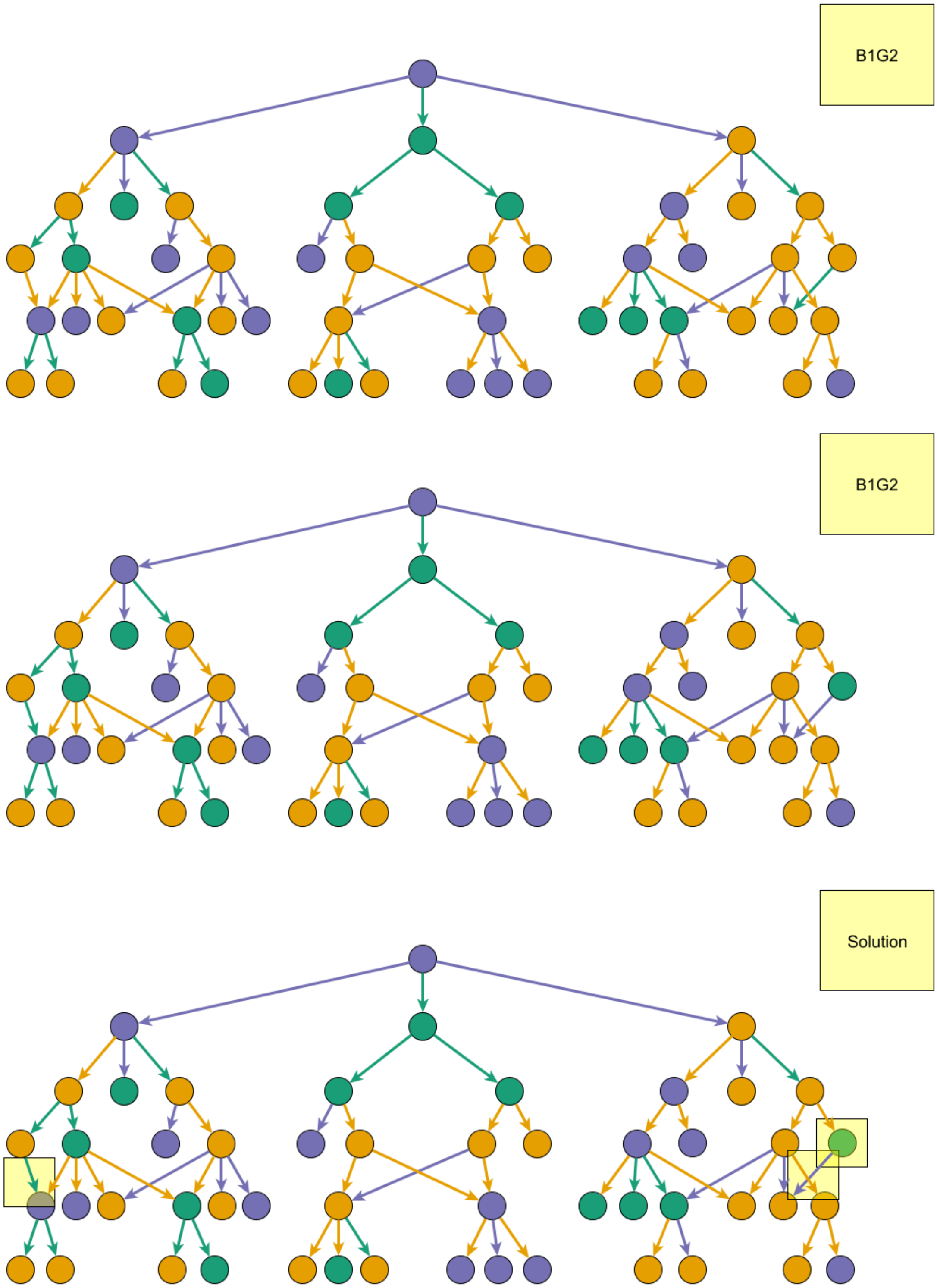

Figure B.2: B1G2

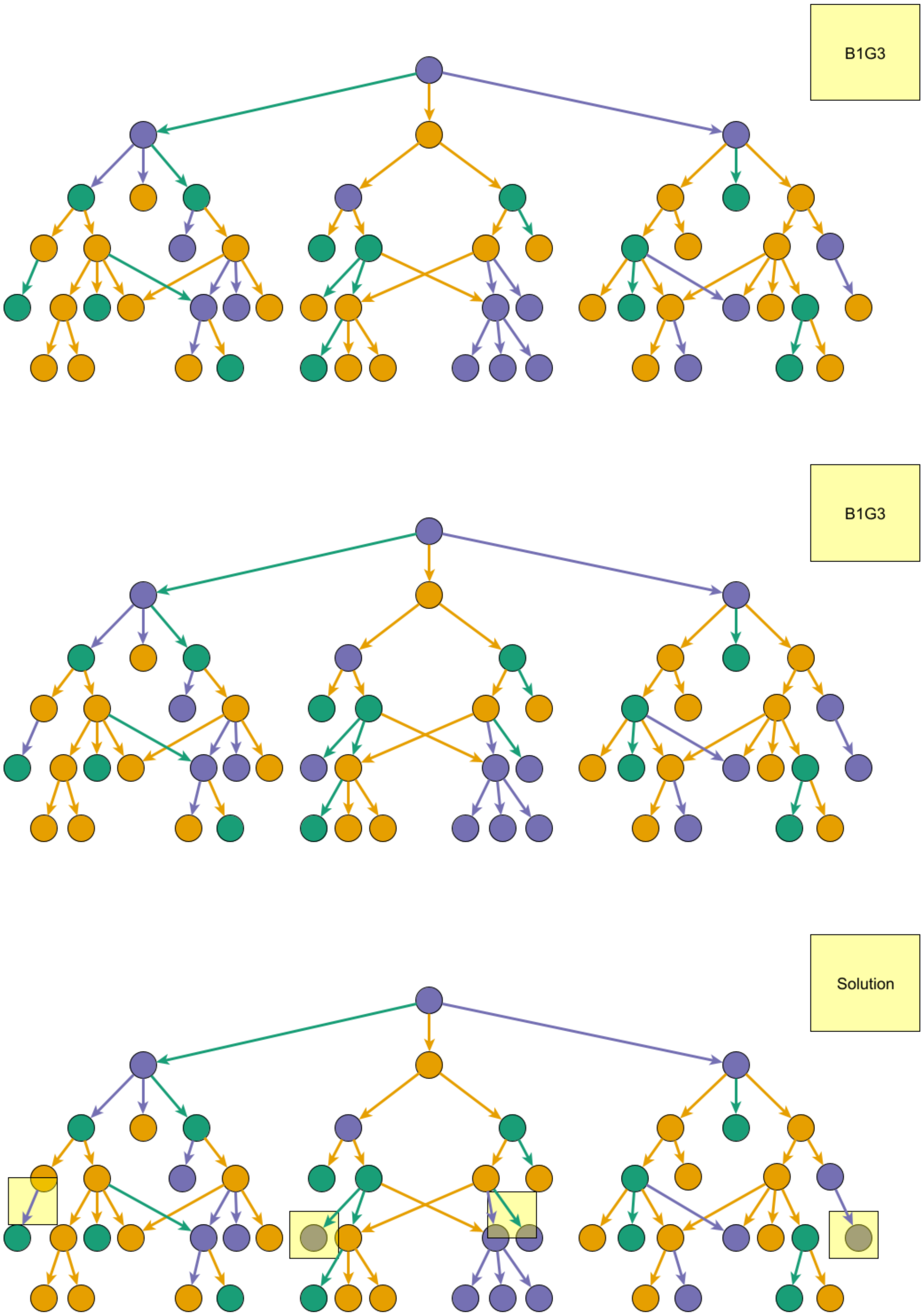

Figure B.3: B1G3

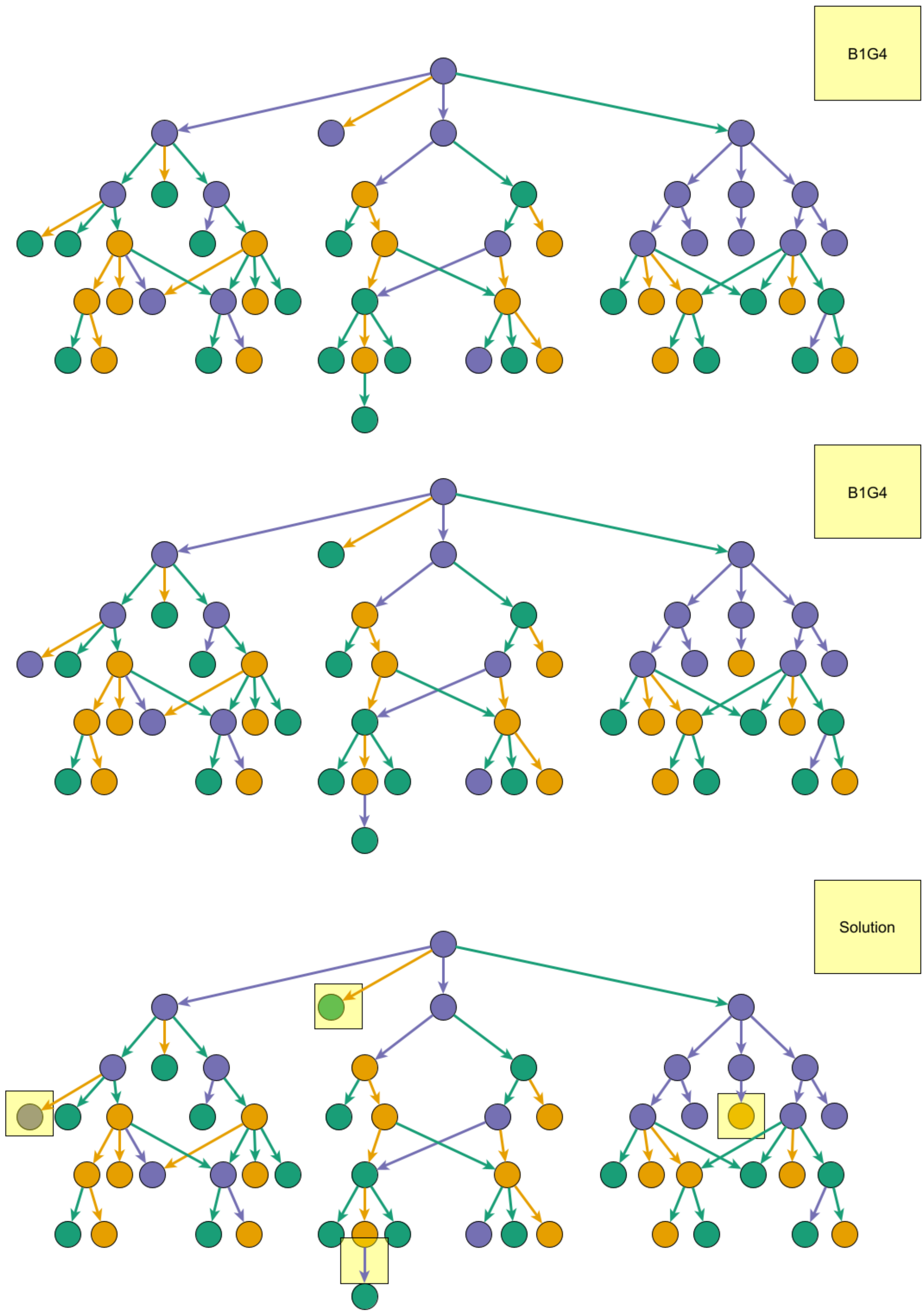

Figure B.4: B1G4

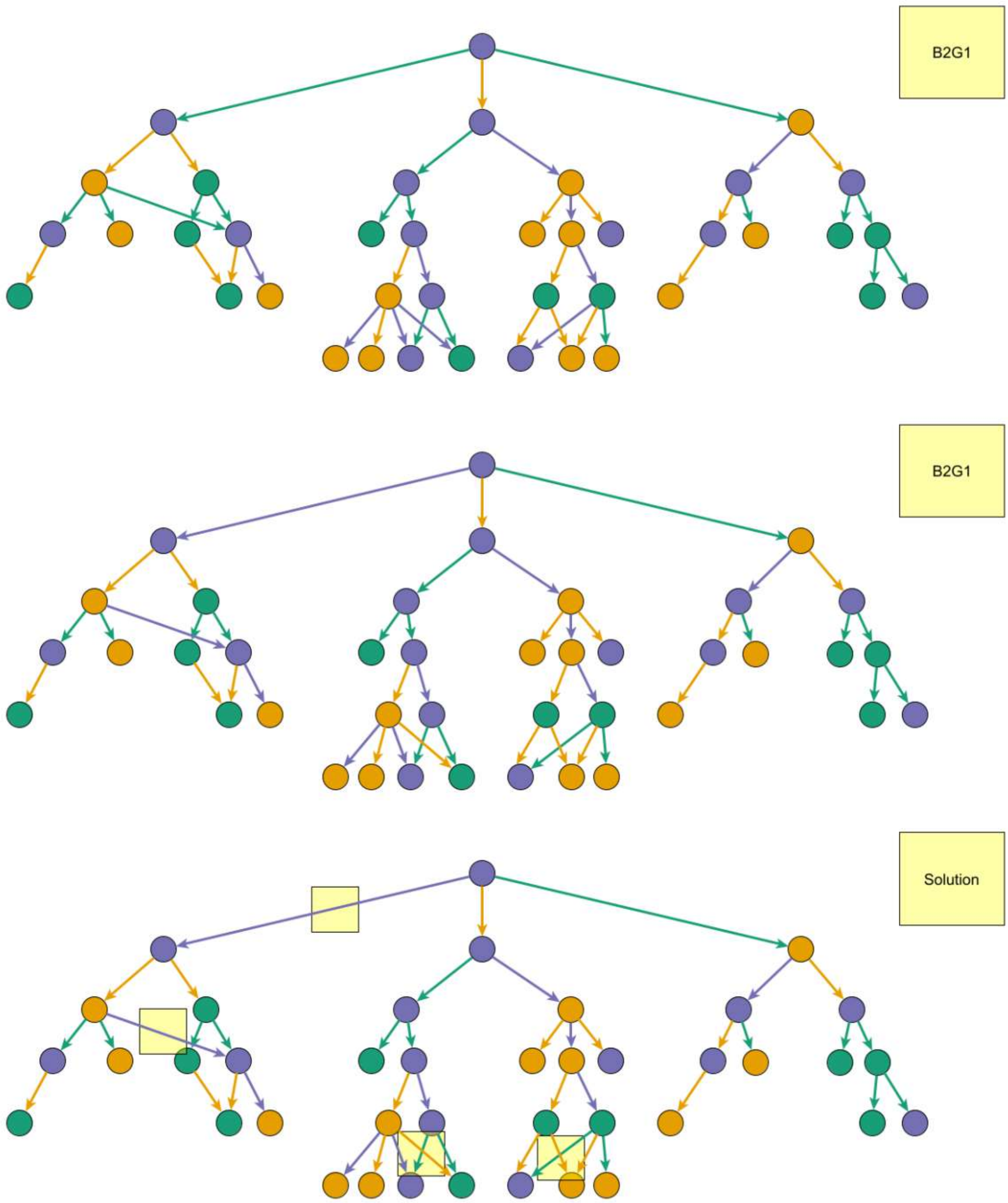

Figure B.5: B2G1

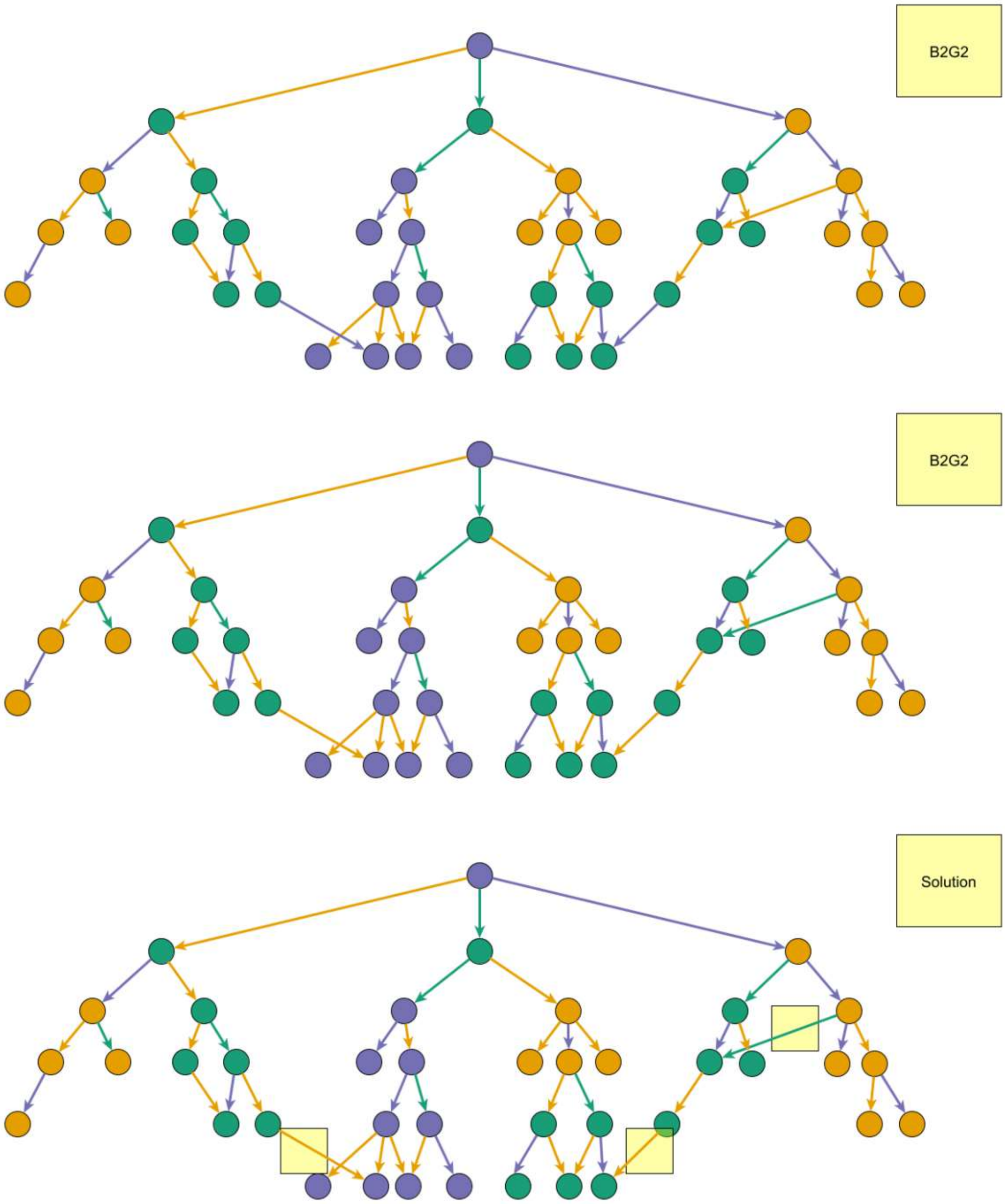

Figure B.6: B2G2

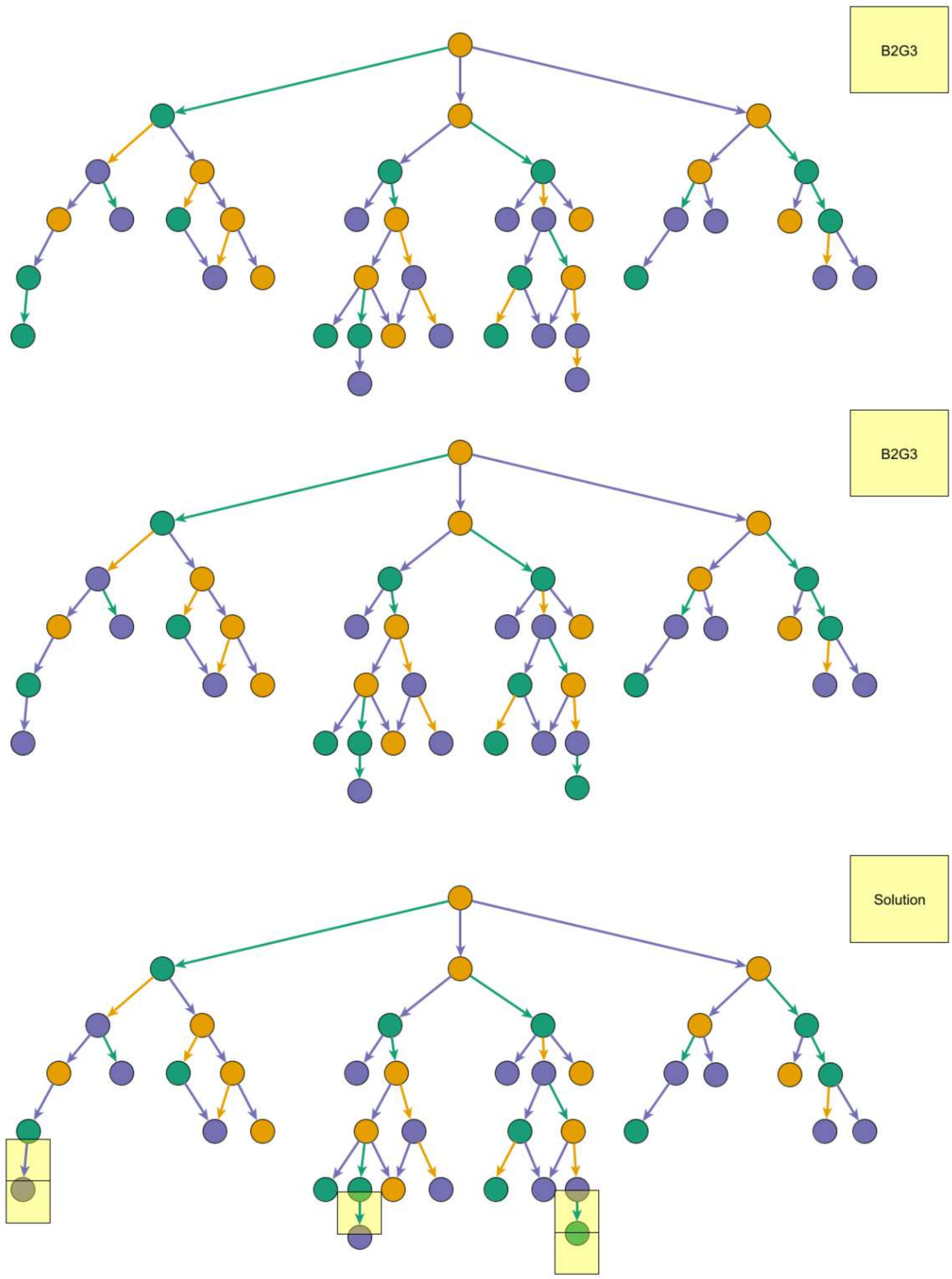

Figure B.7: B2G3

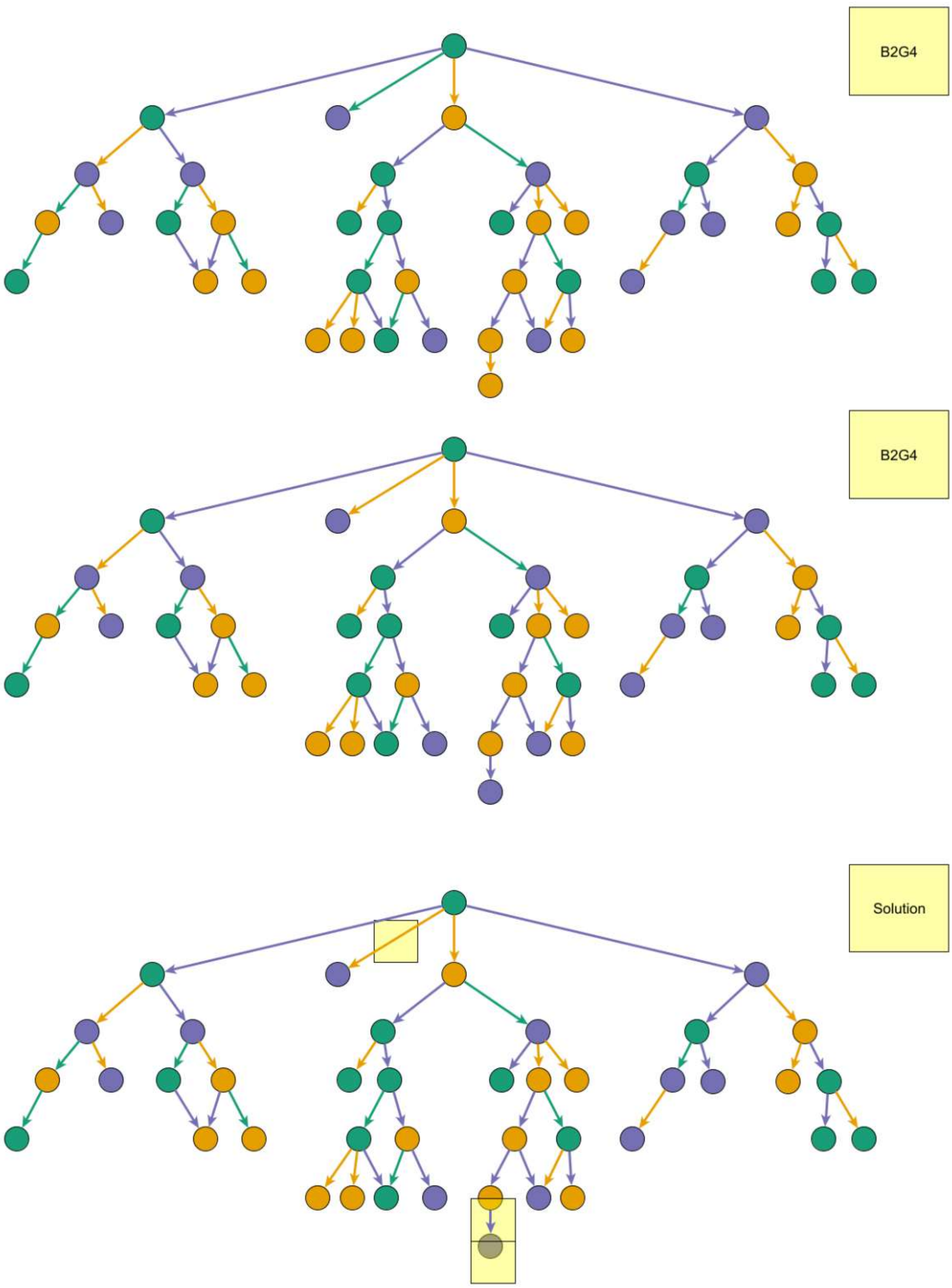

Figure B.8: B2G4

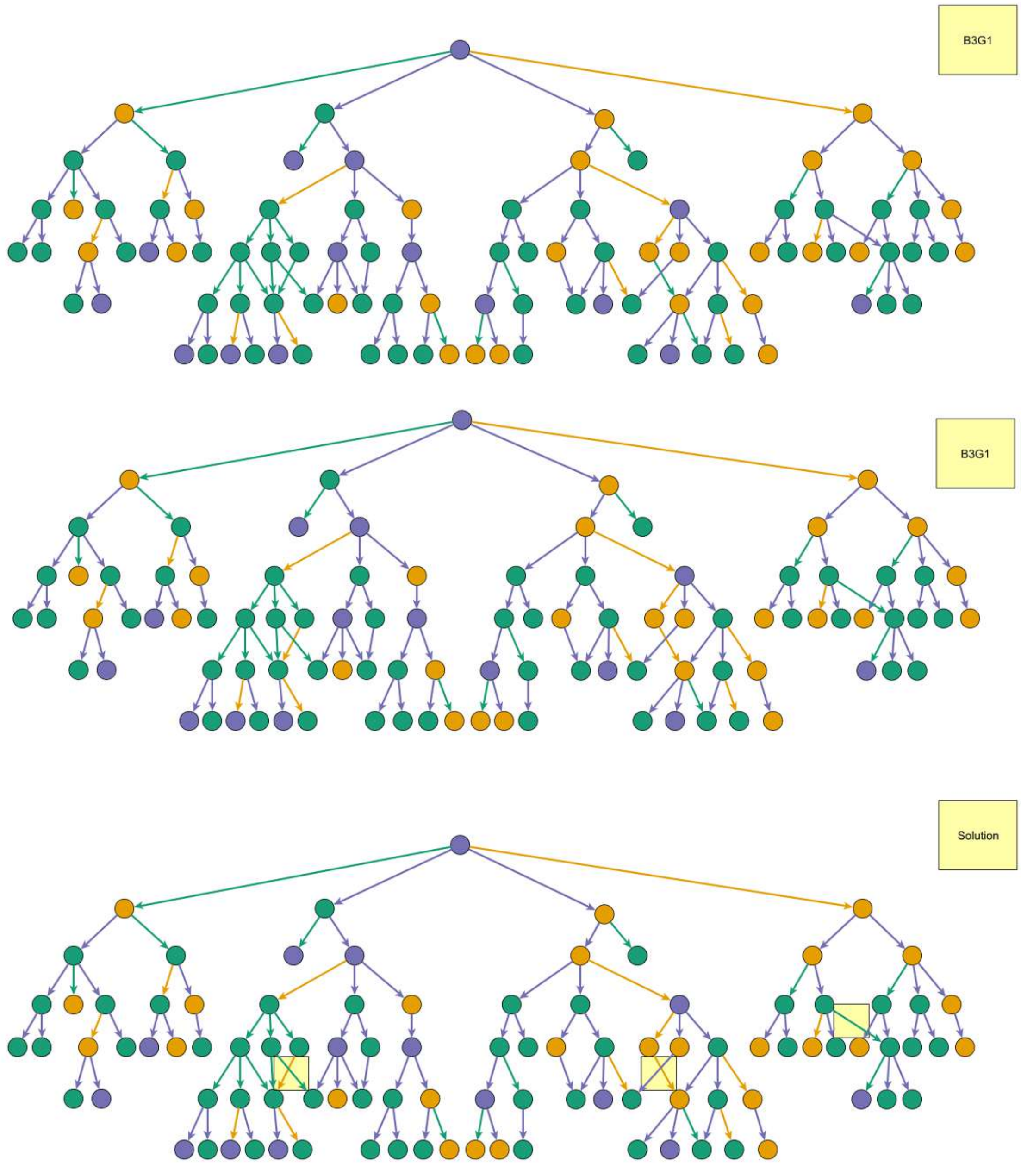

Figure B.9: B3G1

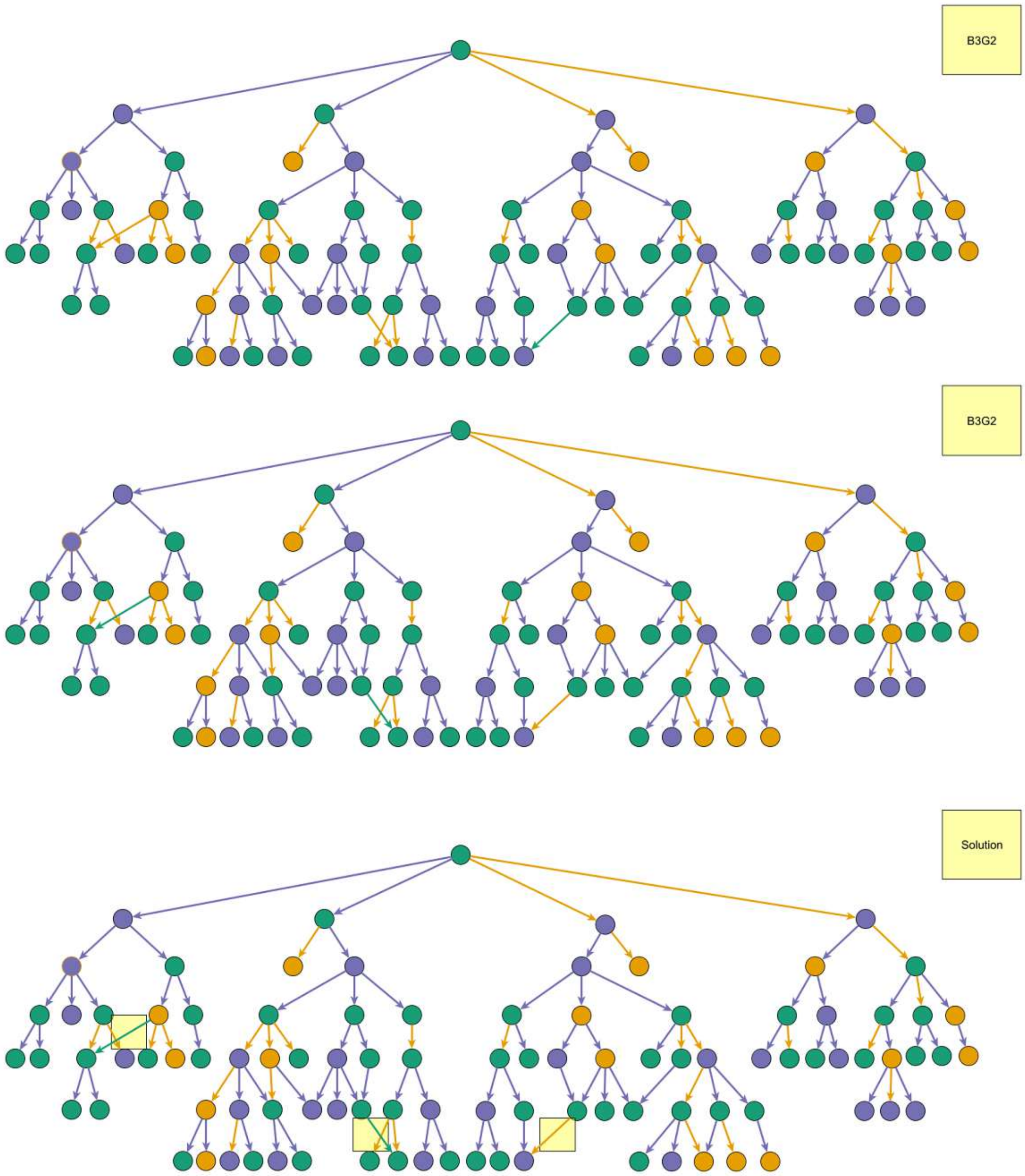

Figure B.10: B3G2

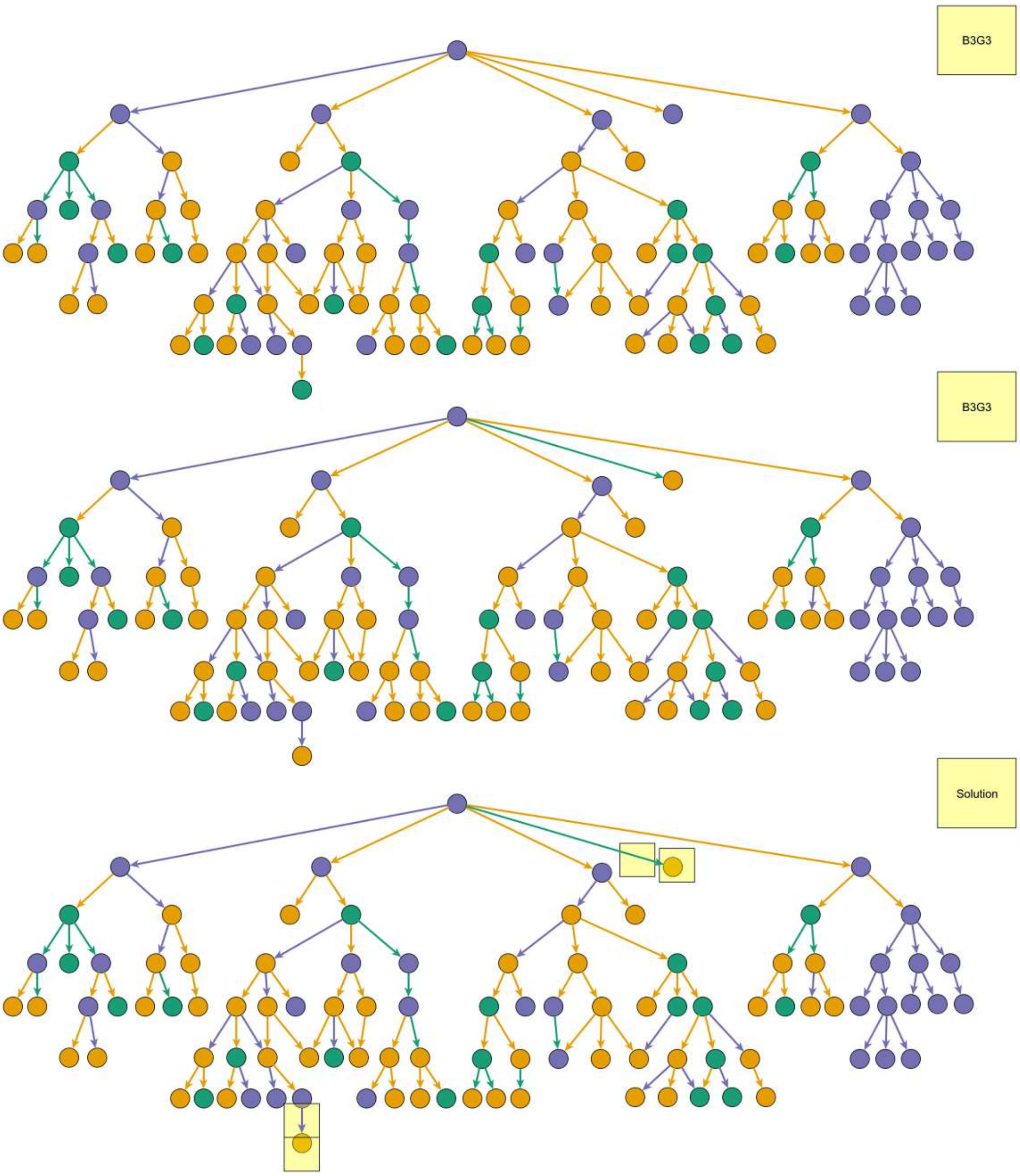

Figure B.11: B3G3

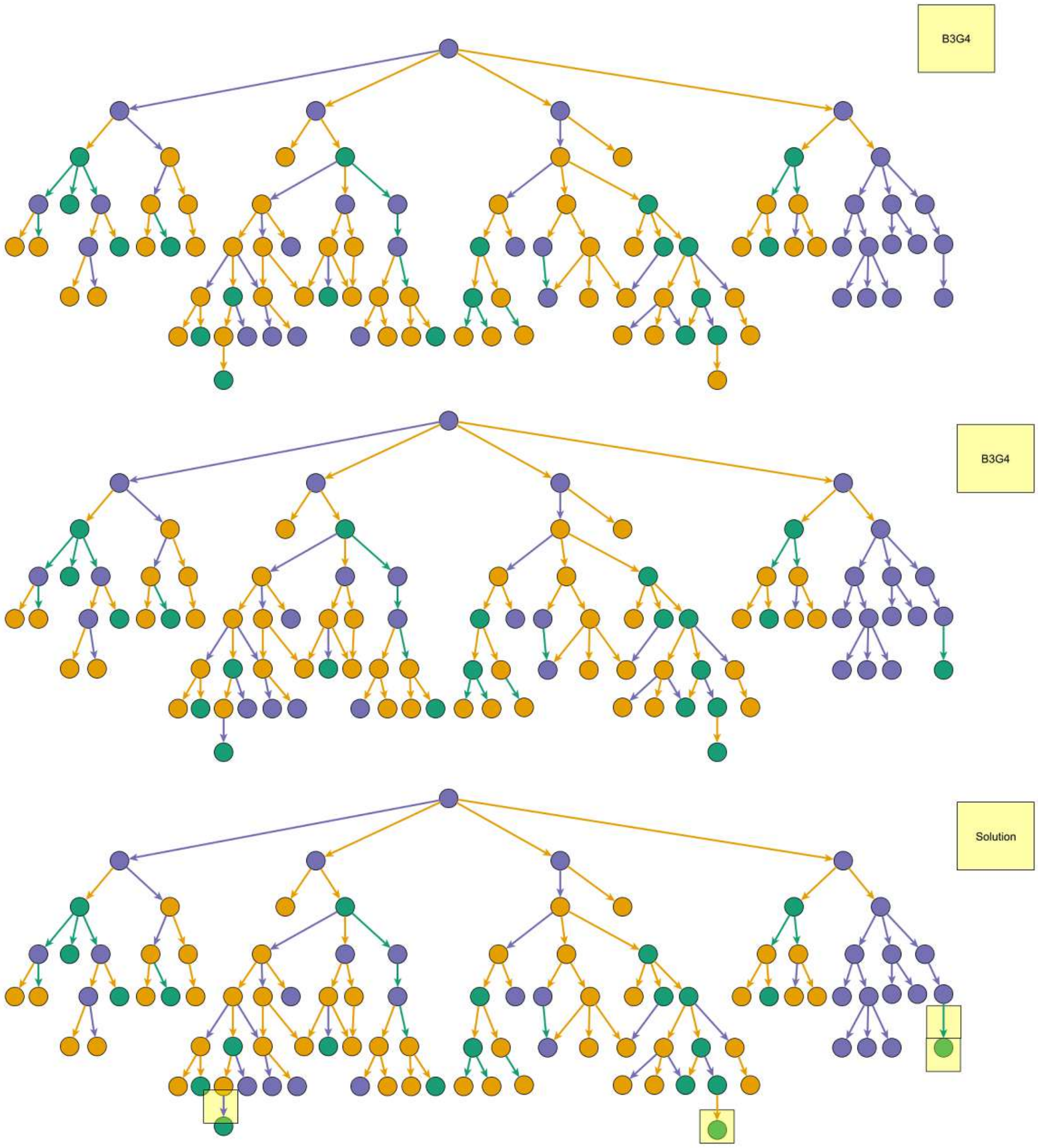

Figure B.12: B3G4

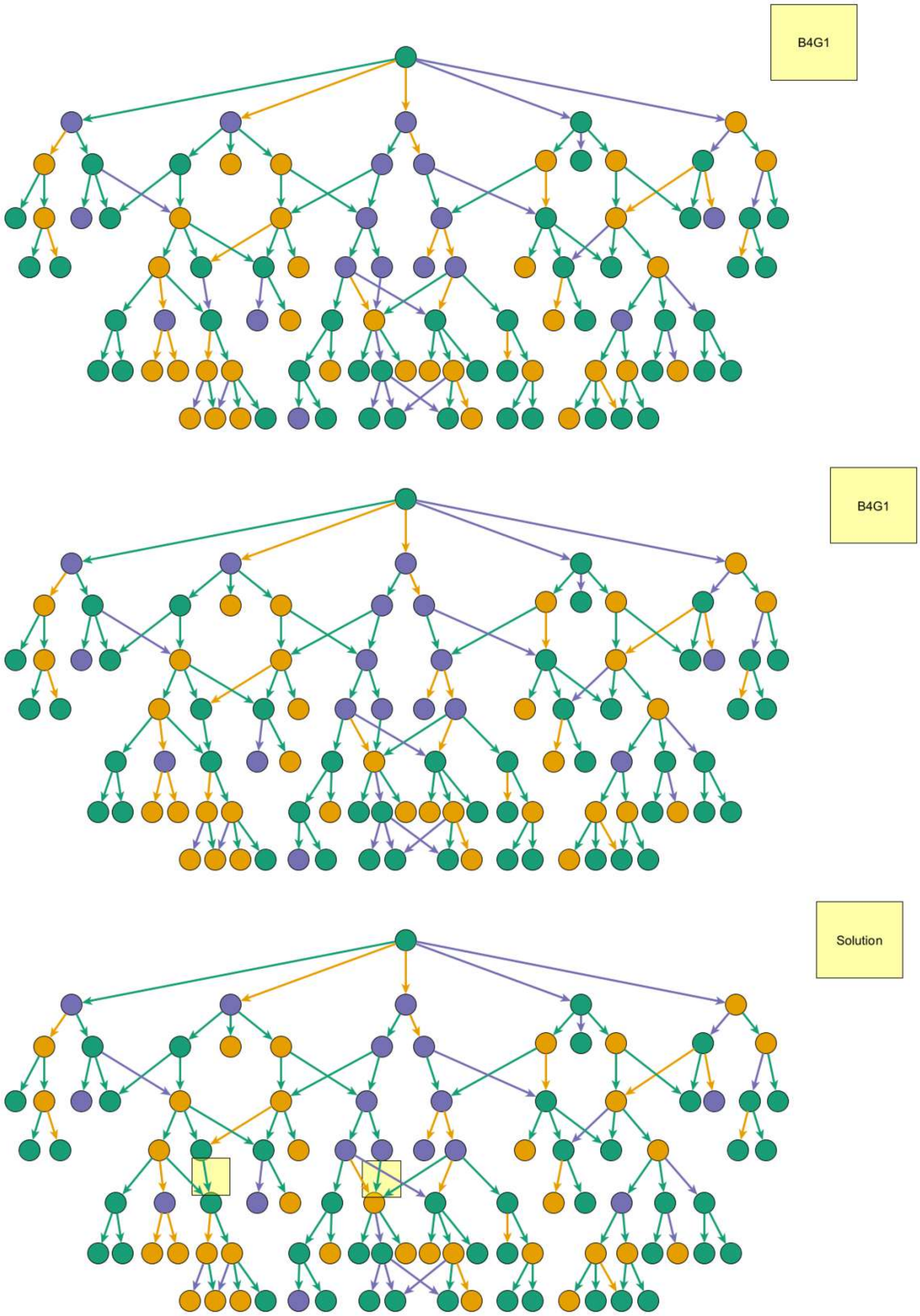

Figure B.13: B4G1

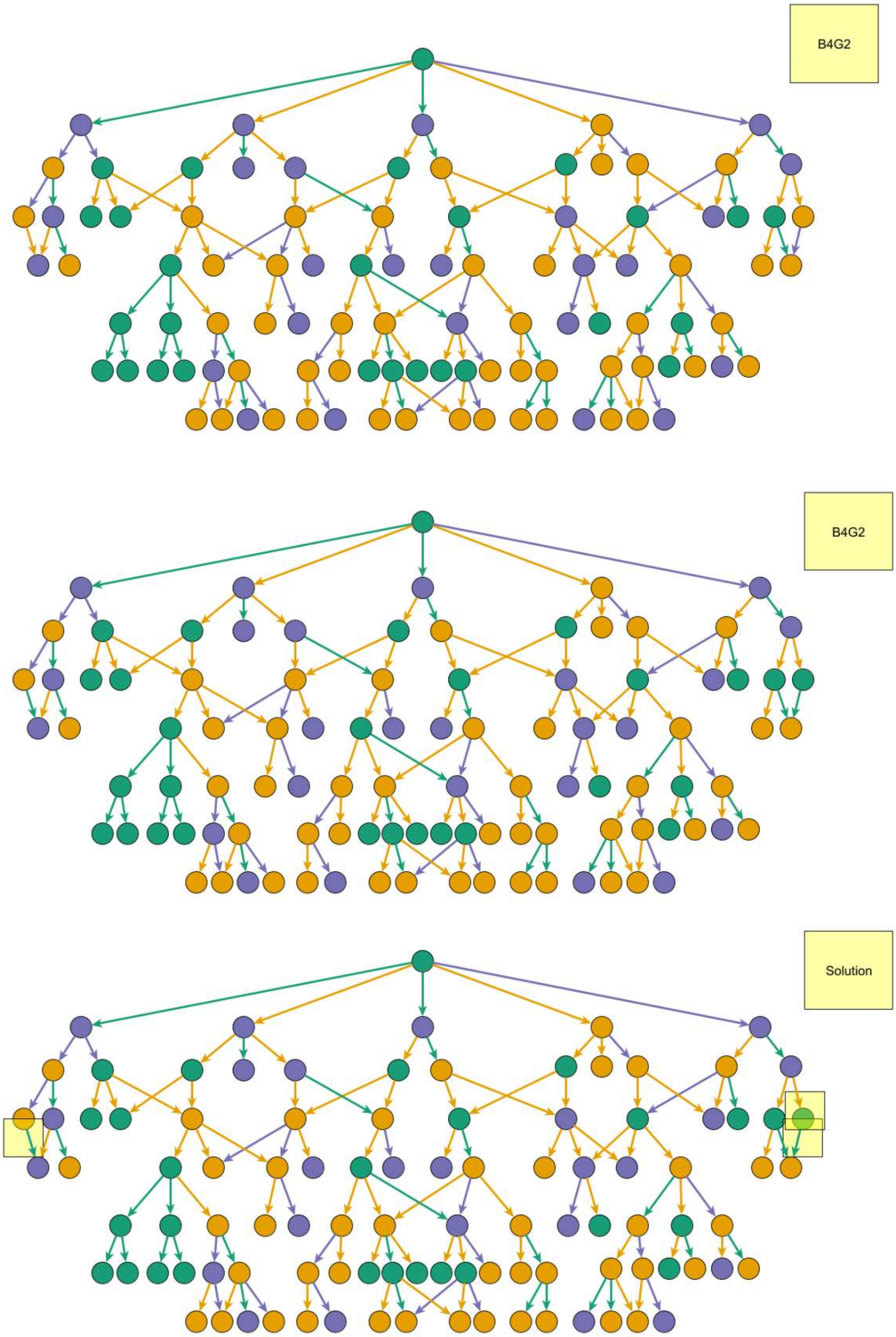

Figure B.14: B4G2

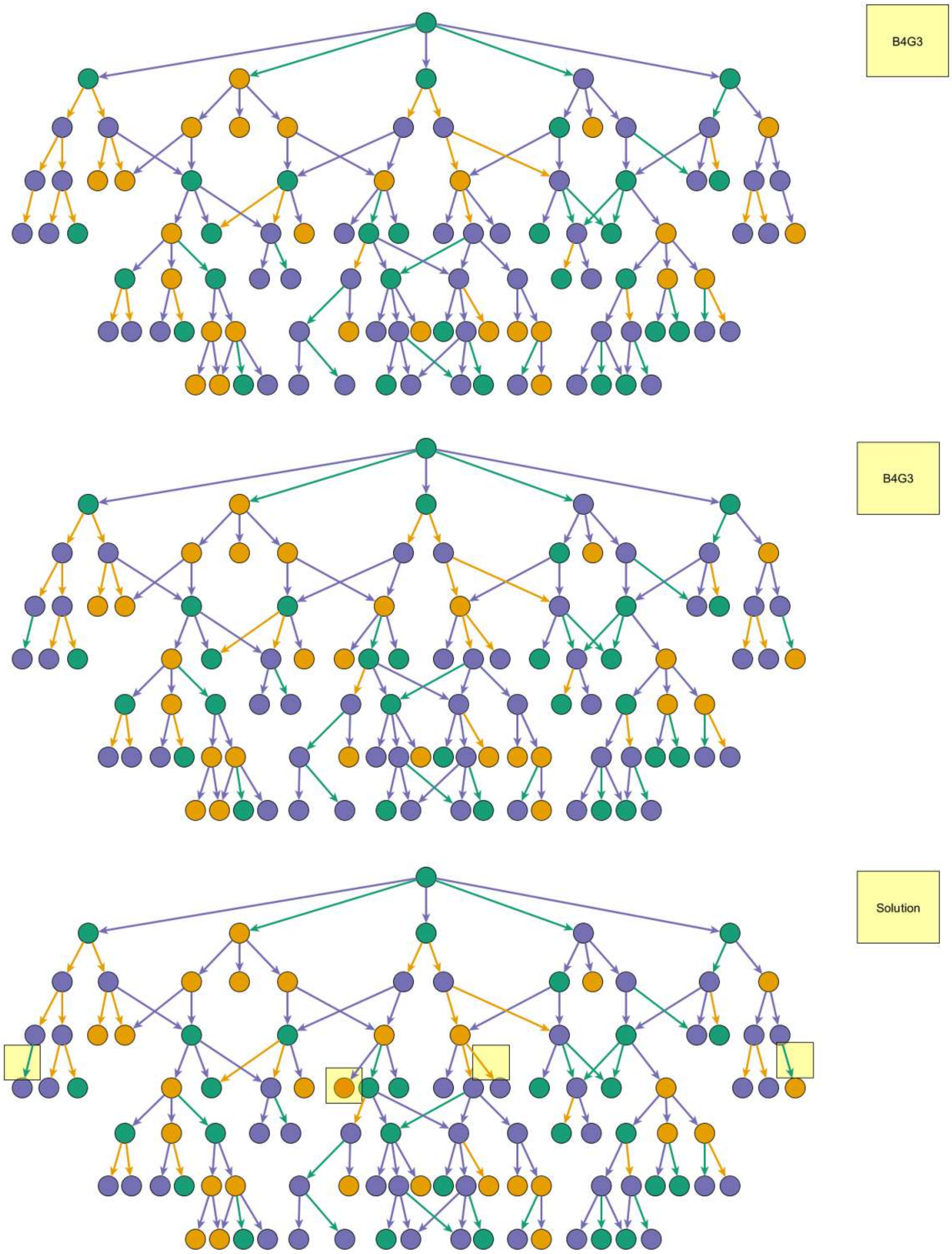

Figure B.15: B4G3

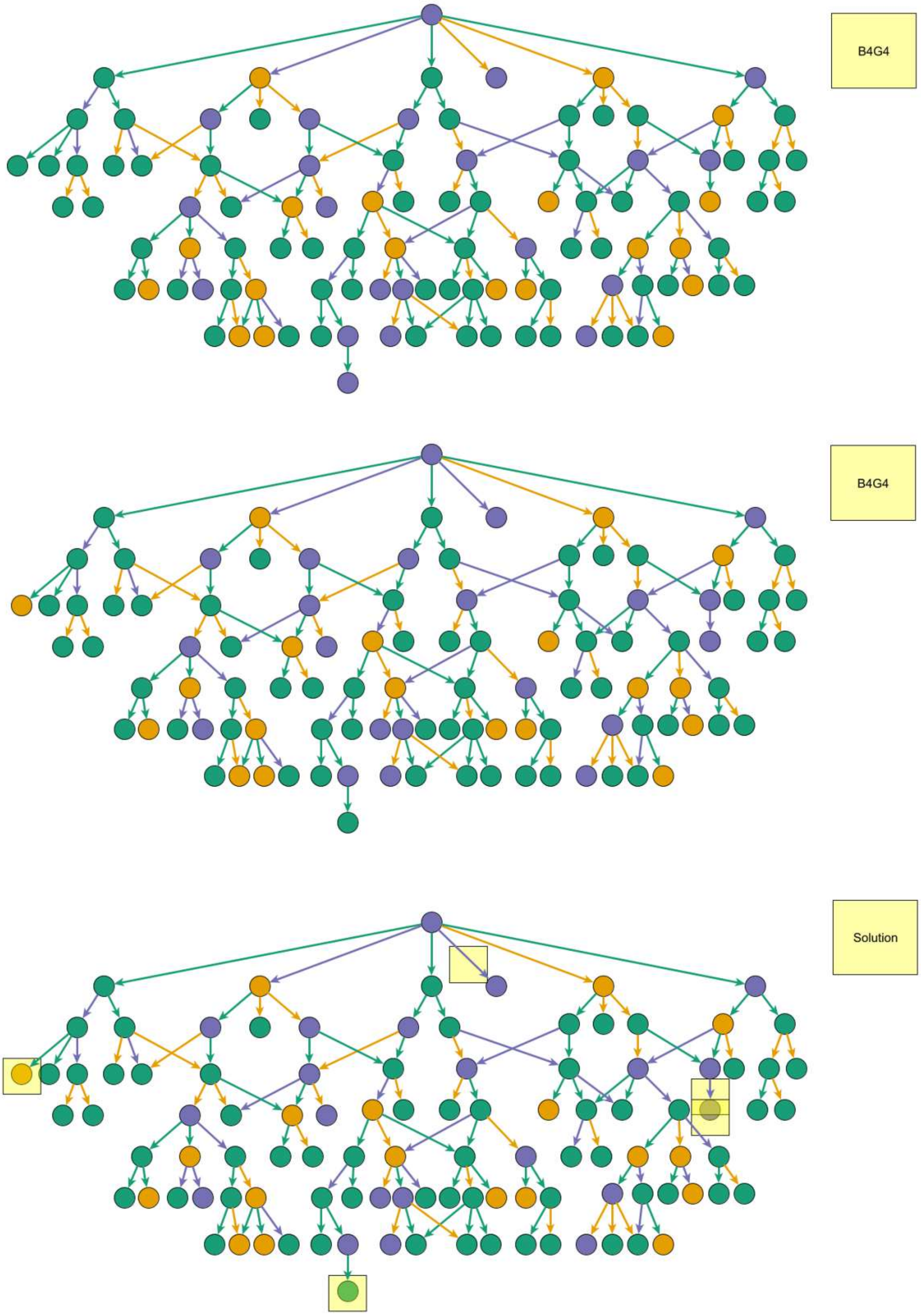

Figure B.16: B4G4

List of Figures

List of Tables

Bibliography

[1] *About Colour Blindness*. en-GB. URL: https://www.colourblindawareness.org/colour-blindness/ (visited on 07/05/2022).

[2] Ahmad Al Badawi and Ali Shatnawi. "Static scheduling of directed acyclic data flow graphs onto multiprocessors using particle swarm optimization". en. In: *Computers & Operations Research* 40.10 (Oct. 2013), pp. 2322–2328. ISSN: 0305-0548. DOI: 10.1016/j.cor.2013.03.015. URL: https://www.sciencedirect.com/science/article/pii/S030505481300097X (visited on 07/20/2022).

[3] Keith Andrews, Martin Wohlfahrt, and Gerhard Wurzinger. "Visual Graph Comparison". In: *2009 13th International Conference Information Visualisation*. ISSN: 2375-0138. July 2009, pp. 62–67. DOI: 10.1109/IV.2009.108.

[4] Kathrin Ballweg et al. "Visual Similarity Perception of Directed Acyclic Graphs: A Study on Influencing Factors". In: *Lecture Notes in Computer Science* 22 (Sept. 2017). DOI: 10.7155/jgaa.00467.

[5] T. Boren and J. Ramey. "Thinking aloud: reconciling theory and practice". en. In: *IEEE Transactions on Professional Communication* 43.3 (Sept. 2000), pp. 261–278. ISSN: 03611434. DOI: 10.1109/47.867942. URL: http://ieeexplore.ieee.org/document/867942/ (visited on 07/05/2022).

[6] Stina Bridgeman and Roberto Tamassia. "A User Study in Similarity Measures for Graph Drawing". en. In: *Graph Drawing*. Ed. by Joe Marks. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2001, pp. 19–30. ISBN: 978-3-540-44541-8. DOI: 10.1007/3-540-44541-2_3.

[7] Peter Buneman et al. "Data citation and the citation graph". en. In: *Quantitative Science Studies* 2.4 (Dec. 2021), pp. 1399–1422. ISSN: 2641-3337. DOI: 10.1162/qss_a_00166. URL: https://direct.mit.edu/qss/article/2/4/1399/108050/Data-citation-and-the-citation-graph (visited on 07/19/2022).

[8] *Chi-Square Test of Independence*. en-AU. URL: https://www.jmp.com/en_au/statistics-knowledge-portal/chi-square-test/chi-square-test-of-independence.html (visited on 07/05/2022).

[9] *Color Difference Calculator*. URL: http://www.brucelindbloom.com/index.html?ColorDifferenceCalc.html (visited on 07/07/2022).

[10] *Color Hex Color Codes*. URL: https://www.color-hex.com/ (visited on 07/07/2022).

[11] *ColorBrewer: Color Advice for Maps*. URL: https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3 (visited on 07/05/2022).

[12] John W. Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. en. Google-Books-ID: 4uB76IC_pOQC. SAGE, 2014. ISBN: 978-1-4522-2610-1.

[13] Patricia Cronin, Frances Ryan, and Michael Coughlan. "Undertaking a literature review: a step-by-step approach". eng. In: *British Journal of Nursing (Mark Allen Publishing)* 17.1 (Jan. 2008), pp. 38–43. ISSN: 0966-0461. DOI: 10.12968/bjon.2008.17.1.28059.

[14] Michela Del Vicario et al. "The spreading of misinformation online". en. In: *Proceedings of the National Academy of Sciences* 113.3 (Jan. 2016), pp. 554–559. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1517441113. URL: https://pnas.org/doi/full/10.1073/pnas.1517441113 (visited on 07/19/2022).

[15] C. Dinesh. "Analysis of the Applicability of Different Color Difference Equations in Printing Industry". en. In: *International Journal of Science, Engineering & Computer Technology* 8.2-4 (2018), pp. 74–79.

[16] K. Anders Ericsson and Herbert A. Simon. *Protocol Analysis: Verbal Reports as Data*. en. Apr. 1993. DOI: 10.7551/mitpress/5657.001.0001. URL: https://direct.mit.edu/books/book/4763/Protocol-AnalysisVerbal-Reports-as-Data (visited on 07/06/2022).

[17] *Fig. 1. A directed acyclic graph G = (V, E). The set of nodes V = {v1,...* en. URL: https://www.researchgate.net/figure/A-directed-acyclic-graph-G-V-E-The-set-of-nodes-V-v1-v2-v15-Note_fig5_276851762 (visited on 07/17/2022).

[18] Jim Frost. *Spearman's Correlation Explained*. en-US. Mar. 2021. URL: https://statisticsbyjim.com/basics/spearmans-correlation/ (visited on 08/01/2022).

[19] *GIMP*. en. URL: https://www.gimp.org/ (visited on 07/05/2022).

[20] Michael Gleicher et al. "Visual comparison for information visualization". en. In: *Information Visualization* 10.4 (Oct. 2011). Publisher: SAGE Publications, pp. 289–309. ISSN: 1473-8716. DOI: 10.1177/1473871611416549. URL: https://doi.org/10.1177/1473871611416549 (visited on 03/21/2021).

[21] Robert L. Goldstone and Ji Yun Son. "Similarity". In: *The Cambridge handbook of thinking and reasoning*. New York, NY, US: Cambridge University Press, 2005, pp. 13–36. ISBN: 978-0-521-82417-0 978-0-521-53101-6.

[22] Maaike van den Haak, Menno De Jong, and Peter Jan Schellens. “Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue”. en. In: *Behaviour & Information Technology* 22.5 (Sept. 2003), pp. 339–351. ISSN: 0144-929X, 1362-3001. DOI: 10.1080/0044929031000. URL: http://www.tandfonline.com/doi/abs/10.1080/0044929031000 (visited on 07/05/2022).

[23] David Harper and Andrew Robert Thompson. *Qualitative research methods in mental health and psychotherapy: a guide for students and practitioners*. eng. Chichester, West Sussex: John Wiley & Sons, 2012. ISBN: 978-1-119-97324-9.

[24] Takako Hashimoto, Tetsuji Kuboyama, and Yukari Shirota. “Rumor analysis framework in social media”. In: *TENCON 2011 - 2011 IEEE Region 10 Conference*. ISSN: 2159-3450. Nov. 2011, pp. 133–137. DOI: 10.1109/TENCON.2011.6129078.

[25] Wanja A. Hemmerich. *Binomiale Logistische Regression: Variablen interpretieren | StatistikGuru.de*. de-DE. URL: https://statistikguru.de/spss/binomiale-logistische-regression/variablen-interpretieren.html (visited on 07/05/2022).

[26] Morten Hertzum, Pia Borlund, and Kristina B. Kristoffersen. “What Do Thinking-Aloud Participants Say? A Comparison of Moderated and Unmoderated Usability Sessions”. en. In: *International Journal of Human-Computer Interaction* 31.9 (Sept. 2015), pp. 557–570. ISSN: 1044-7318, 1532-7590. DOI: 10.1080/10447318.2015.1065691. URL: http://www.tandfonline.com/doi/full/10.1080/10447318.2015.1065691 (visited on 07/05/2022).

[27] Shinobu Ishihara. *Tests for colorblindness*. Tokyo, Japan: Kanehara Shuppan Co, 1972.

[28] K. Kanagaraj and S. Swamynathan. “A realistic approach for representing and scheduling workflows in cloud computing environment”. In: *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Sept. 2016, pp. 1615–1621. DOI: 10.1109/ICACCI.2016.7732279.

[29] Simone Kriglstein and Margit Pohl. “Choosing the Right Sample? Experiences of Selecting Participants for Visualization Evaluation”. In: May 2015. DOI: 10.2312/eurorv3.20151146.

[30] Tatiana von Landesberger et al. “Visual analysis of contagion in networks”. en. In: *Information Visualization* 14.2 (Apr. 2015). Publisher: SAGE Publications, pp. 93–110. ISSN: 1473-8716. DOI: 10.1177/1473871613487087. URL: https://doi.org/10.1177/1473871613487087 (visited on 03/21/2021).

[31] M. Lanzenberger and J. Sampson. “AlViz - A Tool for Visual Ontology Alignment”. en. In: *Tenth International Conference on Information Visualisation (IV’06)*. London, England: IEEE, 2006, pp. 430–440. ISBN: 978-0-7695-2602-7. DOI: 10.1109/IV.2006.18. URL: http://ieeexplore.ieee.org/document/1648296/ (visited on 08/18/2022).

[32] Eric Lehman, Frank Thomson Leighton, and Albert R Meyer. *Mathematics for computer science*. English. OCLC: 1137555943. 2017. ISBN: 978-1-68092-122-9.

[33] Jiafan Li, Yuhua Liu, and Changbo Wang. "Evaluation of graph layout methods based on visual perception". en. In: *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing - ICVGIP '16*. Guwahati, Assam, India: ACM Press, 2016, pp. 1–7. ISBN: 978-1-4503-4753-2. DOI: 10.1145/3009977.3010070. URL: http://dl.acm.org/citation.cfm?doid=3009977.3010070 (visited on 07/04/2022).

[34] Kim Marriott et al. "Memorability of Visual Features in Network Diagrams". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (Dec. 2012). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 2477–2485. ISSN: 1941-0506. DOI: 10.1109/TVCG.2012.245.

[35] V. Minichiello. *In-depth Interviewing: Researching People*. Longman Cheshire, 1990. ISBN: 978-0-582-71272-0. URL: https://books.google.at/books?id=64L3AQAACAAJ.

[36] Haradhan Kumar Mohajan. "QUALITATIVE RESEARCH METHODOLOGY IN SOCIAL SCIENCES AND RELATED SUBJECTS". en. In: *Journal of Economic Development, Environment and People* 7.1 (Mar. 2018), p. 23. ISSN: 2285-3642. DOI: 10.26458/jedep.v7i1.571. URL: http://ojs.spiruharet.ro/index.php/jedep/article/view/jedep.v7i1.571 (visited on 07/04/2022).

[37] Denish Moorthy et al. *Details on the Construction of Citation Graphs*. en. Publication Title: Concordance Between the Findings of Epidemiological Studies and Randomized Trials in Nutrition: An Empirical Evaluation and Citation Analysis: Nutritional Research Series, Vol. 6 [Internet]. Agency for Healthcare Research and Quality (US), May 2013. URL: https://www.ncbi.nlm.nih.gov/books/NBK138254/ (visited on 07/19/2022).

[38] Bernard M. E. Moret et al. "Phylogenetic networks: modeling, reconstructibility, and accuracy". In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2004). DOI: 10.1109/TCBB.2004.10.

[39] Tamara Munzner et al. "TreeJuxtaposer: scalable tree comparison using Focus+Context with guaranteed visibility". en. In: *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03*. San Diego, California: ACM Press, 2003, p. 453. ISBN: 978-1-58113-709-5. DOI: 10.1145/1201775.882291. URL: http://portal.acm.org/citation.cfm?doid=1201775.882291 (visited on 08/18/2022).

[40] Kevin D. O'Gorman and Robert MacIntosh. *Research Methods for Business and Management: a guide to writing your dissertation*. en. Google-Books-ID: IMRLEAAAQBAJ. Goodfellow Publishers Ltd, Sept. 2015. ISBN: 978-1-910158-52-4.

[41] Wendy Olsen. "Triangulation in Social Research: Qualitative and Quantitative Methods Can Really Be Mixed". en. In: (), p. 30.

[42] Anshul Vikram Pandey et al. “Towards Understanding Human Similarity Perception in the Analysis of Large Sets of Scatter Plots”. en. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, May 2016, pp. 3659–3669. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858155. URL: https://dl.acm.org/doi/10.1145/2858036.2858155 (visited on 07/04/2022).

[43] André Queirós, Daniel Faria, and Fernando Almeida. “Strengths And Limitations Of Qualitative And Quantitative Research Methods”. en. In: (Sept. 2017). Publisher: Zenodo. DOI: 10.5281/ZENODO.887089. URL: https://zenodo.org/record/887089 (visited on 07/04/2022).

[44] Md Shidur Rahman. “The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language “Testing and Assessment” Research: A Literature Review”. en. In: *Journal of Education and Learning* 6.1 (Nov. 2016), p. 102. ISSN: 1927-5269, 1927-5250. DOI: 10.5539/jel.v6n1p102. URL: http://www.ccsenet.org/journal/index.php/jel/article/view/64330 (visited on 07/04/2022).

[45] Daniel Reisberg. *Cognition: exploring the science of the mind*. 4th ed. New York: W. W. Norton, 2010. ISBN: 978-0-393-93048-1.

[46] Carsten Schmitz. *STARTSEITE*. de-de. URL: https://limesurvey.org/de/ (visited on 07/05/2022).

[47] Robert Sedgewick and Kevin Daniel Wayne. *Algorithms*. 4th ed. Upper Saddle River, NJ: Addison-Wesley, 2011. ISBN: 978-0-321-57351-3.

[48] Hannah Snyder. “Literature review as a research methodology: An overview and guidelines”. en. In: *Journal of Business Research* 104 (Nov. 2019), pp. 333–339. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2019.07.039. URL: https://www.sciencedirect.com/science/article/pii/S0148296319304564 (visited on 07/17/2022).

[49] S Sofaer. “Qualitative methods: what are they and why use them?” In: *Health Services Research* 34.5 Pt 2 (Dec. 1999), pp. 1101–1118. ISSN: 0017-9124. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089055/ (visited on 07/04/2022).

[50] Paul Solomon. “The think aloud method: A practical guide to modelling cognitive processes”. en. In: *Information Processing & Management* 31.6 (Nov. 1995), pp. 906–907. ISSN: 03064573. DOI: 10.1016/0306-4573(95)90031-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0306457395900314 (visited on 07/05/2022).

[51] Utkarsh Soni et al. “The Perception of Graph Properties in Graph Layouts”. In: *Computer Graphics Forum* 37 (June 2018), pp. 169–181. DOI: 10.1111/cgf.13410.

[52] Richard A. Swanson and Ed Holton, eds. *Research in organizations: foundations and methods of inquiry*. en. 1st ed. San Francisco, CA: Berrett-Koehler Publishers, 2005. ISBN: 978-1-57675-314-9.

[53] Danielle Albers Szafir. "Modeling Color Difference for Visualization Design". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 392–401. ISSN: 1941-0506. DOI: 10.1109/TVCG.2017.2744359.

[54] R. Murray Thomas. *Blending Qualitative and Quantitative Research Methods in Theses and Dissertations*. en. Corwin Press, Mar. 2003. ISBN: 978-0-7619-3932-0.

[55] K. Thulasiraman and M. N. S. Swamy. *Graphs: theory and algorithms*. New York: Wiley, 1992. ISBN: 978-0-471-51356-8.

[56] Amos Tversky. "Features of similarity." en. In: *Psychological Review* 84.4 (1977), pp. 327–352. ISSN: 0033-295X. DOI: 10.1037/0033-295X.84.4.327. URL: http://content.apa.org/journals/rev/84/4/327 (visited on 07/20/2022).

[57] Guenter Wallner et al. "Perception of Differences in Directed Acyclic Graphs: Influence Factors & Cognitive Strategies". In: *Proceedings of the 31st European Conference on Cognitive Ergonomics*. ECCE 2019. New York, NY, USA: Association for Computing Machinery, Sept. 2019, pp. 57–64. ISBN: 978-1-4503-7166-7. DOI: 10.1145/3335082.3335083. URL: https://doi.org/10.1145/3335082.3335083 (visited on 03/21/2021).

[58] Günter Wallner et al. "Influence of Shape, Density, and Edge Crossings on the Perception of Graph Differences". In: *Diagrammatic Representation and Inference*. Ed. by Ahti-Veikko Pietarinen et al. Cham: Springer International Publishing, 2020, pp. 348–356. ISBN: 978-3-030-54249-8.

[59] *What is Delta E? And Why Is It Important for Color Accuracy?* en-US. Jan. 2021. URL: https://www.viewsonic.com/library/creative-work/what-is-delta-e-and-why-is-it-important-for-color-accuracy/ (visited on 07/07/2022).

[60] Amy E. Whitehead, Jamie A. Taylor, and Remco C. J. Polman. "Examination of the suitability of collecting in event cognitive processes using Think Aloud protocol in golf". In: *Frontiers in Psychology* 6 (2015). ISSN: 1664-1078. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01083 (visited on 07/06/2022).

[61] yWorks company the diagramming yWorks the diagramming. *yEd Graph Editor*. en. URL: https://www.yworks.com/products/yed (visited on 07/05/2022).

[62] Zach. *How to Interpret glm Output in R (With Example)*. en-US. Nov. 2021. URL: https://www.statology.org/interpret-glm-output-in-r/ (visited on 07/05/2022).

[63] Tingting Zhao, Sharon McDonald, and Helen M. Edwards. "The impact of two different think-aloud instructions in a usability test: a case of just following orders?" en. In: *Behaviour & Information Technology* 33.2 (Feb. 2014), pp. 163–183. ISSN: 0144-929X, 1362-3001. DOI: 10.1080/0144929X.2012.708786. URL: http://www.tandfonline.com/doi/abs/10.1080/0144929X.2012.708786 (visited on 07/06/2022).

[64] Liang Zhou and Charles Hansen. "A Survey of Colormaps in Visualization". In: *IEEE transactions on visualization and computer graphics* 22 (Oct. 2015). DOI: 10.1109/TVCG.2015.2489649.

[65] Nataša Živić, Esad Kadušić, and Kerim Kadušić. "Directed Acyclic Graph as Hashgraph: an Alternative DLT to Blockchains and Tangles". In: *2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH)*. Mar. 2020, pp. 1–4. DOI: 10.1109/INFOTEH48170.2020.9066312.