Effective Visual Analytics for Exploring Argumentation and Deliberation in Text Corpora

by

Dora Kiesel

Dissertation to obtain the academic degree of
Dr. rer. nat.

Faculty of Media
Bauhaus-Universität Weimar
Germany

24. November 2025

1. Referee: Prof. Dr. Bernd Fröhlich
Bauhaus-Universität Weimar
Weimar, Deutschland

2. Referee: Prof. Dr. Andreas Kerren
Linköping University
Linköping, Sweden

Abstract

Effective Visual Analytics for Exploring Argumentation and Deliberation in Text Corpora

Argumentation is a natural part of human communication and a focus of interest for social scientists and linguists. Nevertheless, few visualization and interaction techniques visually support social scientists working with argumentative and deliberative texts. To address this research gap, this thesis investigates new techniques for visually comparing and analysing the argumentative and deliberative texts of a whole corpus in order to facilitate scholarly work and enable exploratory research on such text corpora.

The investigation proceeds from three perspectives:

1. Visual techniques to express both the hierarchical relationships of multiple argumentation structures and the order of their arguments, supporting detailed interactive analysis;

2. Visual approaches for analyzing deliberative processes in multiple hierarchical thread structures of online discussions;

3. Interaction techniques to smoothly transition between scatter plots and parallel coordinates, enabling the exploration of statistical measures of argumentation structures in relation to other factors such as topics and human values.

The work advances visualization research through several contributions: novel aggregation and visualization techniques for tree sets; a scalable rose-chart-like glyph design for representing node-related attributes in aggregated tree sets, along with a related perception study; and a visual computational notebook for interactive, data-driven research. Furthermore, a seamless geometric transition between scatter plots and parallel coordinates facilitates complex analyses, while additional point representatives in parallel coordinates enhance cluster and correlation identification.

These contributions not only advance visualization research but also provide the social sciences with new visual analysis tools that implement the developed techniques to evaluate their practical value. Many of these techniques enable new ways of conducting research – exploratory as well as

hypothesis-driven – offering fresh insights from existing data and a broader range of analysis possibilities for future study results. In teaching, these results help visually explain argumentation structures, deliberative processes, and the relationship between parallel coordinates and scatter plots.

Abstract (in German)

Effective Visual Analytics for Exploring Argumentation and Deliberation in Text Corpora

Argumentation ist ein natürlicher Teil der menschlichen Kommunikation und ein Kernforschungsthema in den Sozialwissenschaften und der Linguistik. Dennoch gibt es nur wenige Visualisierungs- und Interaktionstechniken, die Wissenschaftlerinnen und Wissenschaftler bei der Arbeit mit argumentativen und deliberativen Texten unterstützen. Um diese Forschungslücke zu schließen, entwickelt diese Arbeit neue Visualisierungstechniken um alle argumentativen und deliberativen Texte eines Korpus miteinander zu vergleichen und zu analysieren, um die wissenschaftliche Arbeit zu erleichtern und die explorative Erforschung der Korpora zu ermöglichen.

Dabei wurden drei Perspektiven betrachtet:

1. Visualisierungstechniken, die die hierarchischen Beziehungen mehrerer Argumentationsstrukturen sowie die Reihenfolge der Argumente in den Texten darstellen und eine detaillierte interaktive Analyse dessen ermöglichen können;

2. Visualisierungstechniken zur Analyse der deliberativen Prozesse in hierarchischen Thread-Strukturen von Online-Diskussionen;

3. Interaktionstechniken die den fließenden Übergang zwischen Streudiagrammen und parallelen Koordinaten ermöglichen, um die statistischen Maße der hierarchischen Argumentationsstrukturen zu untersuchen und sie mit anderen Aspekten wie Themen und menschlichen Werten in Beziehung zu setzen.

Die Arbeit liefert folgende Kernbeiträge zur Visualisierungsforschung: neue Visualisierungstechniken für Mengen hierarchischer Strukturen, wie beispielsweise aggregierte Argumentationsbäume oder Thread-Strukturen; ein Entwurf eines skalierbaren Rose-Chart-ähnlichen Glyphen zur Darstellung von Kommentarattributen in aggregierten Thread-Strukturen, sowie eine diesbezügliche Wahrnehmungsstudie; und einem Entwurf eines visuellen Notizbuch für interaktive, datengetriebene Forschungsnotizen. Die

geometrisch fundierte und nahtlose Transition zwischen Streudiagrammen und Parallelen Koordinaten erleichtert komplexe Analysen, während zusätzliche Punktrepräsentanten in den Parallelen Koordinaten Cluster und Korrelationen besser identifizierbar machen.

Nicht nur die Visualisierungsforschung profitiert von der Forschung der Dissertation. Die Sozialwissenschaften erhalten neue visuelle Analysewerkzeuge, die die entwickelten Visualisierungs- und Interaktionstechniken implementieren, um den praktischen Wert der Techniken zu evaluieren. Viele dieser Techniken ermöglichen neue Wege zur Durchführung der Forschungsarbeiten durch Exploration zusätzlich zur Hypothesenprüfung, erbringen neue Erkenntnisse aus bereits bekannten Daten und neue Möglichkeiten der Analyse zukünftiger Forschungsdaten. Die Inhalte der Dissertation bieten als Beitrag für die Lehre auch Mittel, um den Aufbau einer guten Argumentation, deliberative Prozesse und die Beziehung zwischen Parallelen Koordinaten und Streudiagrammen anschaulich zu erklären.

Publications

The thesis consists of several peer-reviewed scientific publications. Besides these, further research has been conducted in cooperation with researchers of adjacent research areas. Figure 1 provides an overview of the relations between the publications listed.

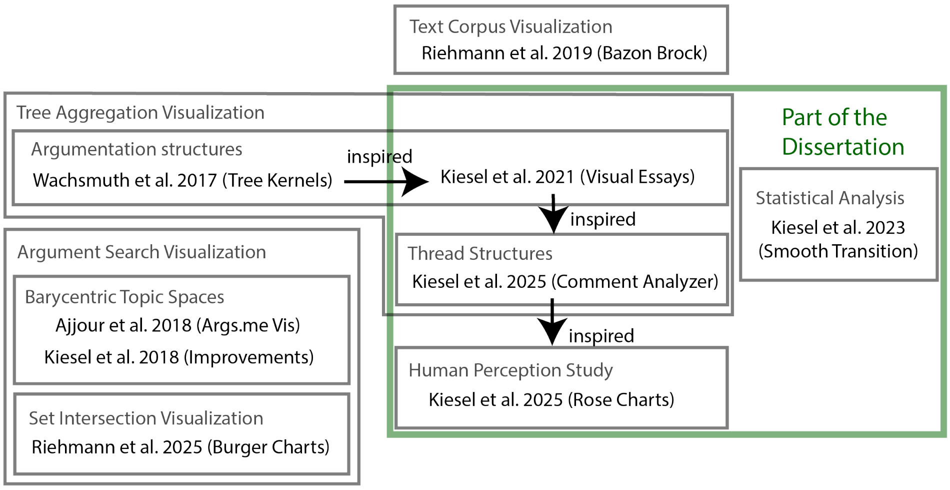


FIGURE 1: Overview of the relations between the publications listed in this section.

Publications presented in this Thesis

D. Kiesel, P. Riehmann, H. Wachsmuth, B. Stein and B. Froehlich, "Visual Analysis of Argumentation in Essays," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 1139-1148, Feb. 2021, doi: 10.1109/TVCG.2020.3030425

D. Kiesel, P. Riehmann, I. Engelmann, H. Ramezani and B. Froehlich, "Comment Analyzer: A Tool for Analyzing Comment Sets and Thread Structures of News Articles," accepted for IEEE Transactions on Visualization and Computer Graphics, 2025, doi: 10.1109/TVCG.2025.3544733

D. Kiesel, P. Riehmann, and B. Froehlich, "Rose Charts: Area or Length Encoding for Fill Level of Circle Sectors?," presented at EuroVis 2025 (short paper), 2025

D. Kiesel, P. Riehmann, and B. Froehlich, "Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots,". in Computer Graphics Forum, no. 42, pp. e14923, 2023, doi:10.1111/cgf.14923

Further Research and Co-Authorships

P. Riehmann, D. Kiesel, and B. Froehlich, "Burger Charts: A Quantitative Display of Set Intersections," presented at EuroVis 2025 (short paper), 2025.

P. Riehmann, D. Kiesel, M. Kohlhaas, and B. Froehlich,"Visualizing a Thinker's Life," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 4, pp. 1803-1816, 2019.

D. Kiesel, P. Riehmann, F. Fan, Y. Ajjour, H. Wachsmuth, B. Stein, and B. Fröhlich, "Improving Barycentric Embeddings of Topics Spaces," in IEEE VIS 2018, 2018. IEEE.

Y. Ajjour, H. Wachsmuth, D. Kiesel, P. Riehmann, F. Fan, G. Castiglia, R. Adejoh, B. Froehlich, and B. Stein. "Visualization of the Topic Space of Argument Search Results in args.me." in 23rd Conference on Empirical Methods in Natural Language Processing (EMNLP 2018) – System Demonstrations, pp. 60–65, 2018. Association for Computational Linguistics.

H. Wachsmuth, G. Da San Martino, D. Kiesel, and B. Stein. "The Impact of Modeling Overall Argumentation with Tree Kernels", in 22nd Conference on Empirical Methods in Natural Language Processing (EMNLP 2017), pp. 2369–2379, 2017. Association for Computational Linguistics.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Teile der Arbeit, die bereits Gegenstand von Prüfungsarbeiten waren, sind ebenfalls unmissverständlich gekennzeichnet. Bei Auswahl und Auswertung des folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise geholfen:

Prof. Bernd Fröhlich	Wissenschaftliche Betreuung der gesamten Arbeit.
Dr. Patrick Riehmann	Diskussionen und Feedback; Durchführung von Expert Reviews für das Paper “Visual Analysis of Argumentation in Essays”.
Prof. Ines Engelmann und Hanna Ramezani	Bereitstellung des Deliberationskorpus und domain-spezifischer, wissenschaftlicher Fragestellungen; Teilnahme an mehreren Expert Review Runden zur Entwicklung des “Comment Analyzers.”
Prof. Henning Wachsmuth	Bereitstellung von Forschungsfragen der Computerlinguistik; Teilnahme an Expert Reviews und Diskussionen; initialer Ideengeber für die Baumaggregation.
Prof. Benno Stein	Teilnahme an Expert Reviews / Diskussionen.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Weimar, 13. Juni 2025 ________________________

Dora Kiesel

Acknowledgments

I would like to thank all who supported me in my endeavor to start, develop, and finish this dissertation.

Especially, I would like to thank my scientific supervisor **Prof. Dr. Bernd Fröhlich** for his support through all phases of writing the dissertation. Without his seemingly limitless patience, his scientific focus and his motivating words, the dissertation would not have been completed.

I would also like to thank **Prof. Dr. Andreas Kerren** for his willingness to act as second reviewer for the thesis.

Finally, I would like to thank **Dr. Patrick Riehmann** for the many scientific discussions and the fruitful cooperation.

Contents

1 **Introduction** 1

- 1.1 Excursion into Argumentation Theory . . . 2
- 1.2 Strategies for Developing Visual Analysis Techniques for Argumentation . . . 4
- 1.3 Research Questions and Contributions . . . 7

2 **Visual Analysis of Argumentation in Essays** 13

- 2.1 Introduction . . . 14
- 2.2 Argumentation Model and Corpora . . . 16
- 2.3 Related Work . . . 17
- 2.4 Argument Structure Analysis System . . . 19
- 2.5 Expert Reviews and Domain Findings . . . 30
- 2.6 Discussion . . . 36
- 2.7 Conclusion and Future Work . . . 38

3 **Comment Analyzer** 41

- 3.1 Introduction . . . 42
- 3.2 Related Work . . . 45
- 3.3 Corpus, Requirements and Tasks . . . 46
- 3.4 Design and Evaluation Process . . . 48
- 3.5 Comment Analyzer: Design and Application . . . 49
- 3.6 Domain Findings with Associated Experts . . . 57
- 3.7 Reviews and Case Studies with Independent Experts . . . 62
- 3.8 Scalability and Generalization . . . 65
- 3.9 Conclusion and Future Work . . . 66

4 **Fill Level of Circle Sectors** 69

- 4.1 Introduction . . . 70
- 4.2 Related Work . . . 70
- 4.3 Pilot Study . . . 71
- 4.4 Main study . . . 73
- 4.5 Discussion, Guidelines and Future Work . . . 78

5 **Smooth Transitions Between Parallel Coordinates and Scatter Plots** 81

5.1 Introduction 82
5.2 Related Work 85
5.3 The need for a transition mediator: the Polycurve Star Plot . 87
5.4 Smooth Line Transition between PCP and SP 89
5.5 Smooth Point Transition between PCP and SP 91
5.6 Interactive Analysis Tool 95
5.7 Proof of Concept with Known Datasets 98
5.8 User Study 99
5.9 Scalability and Appearance 101
5.10 Conclusion and Future Work 103
5.11 Extension: Fine-grained Analysis of the AAE2 corpus 103

6 Conclusion **109**
6.1 Main Contributions 109
6.2 Text visualization and LLMs 113
6.3 Open Problems and Future Work 114

References **116**

1 Introduction

Argumentation is key to human communication and writing, and has been subject to research by humanists and linguists for a long time. Aristotle [4] describes the three modes of persuasion along with different *Topoi* – rhetorical moves to build arguments – he observed in public talks, Walton [149] observed and formalized common argumentation schemes, and van Eemeren [142] studies argumentation as complex speech act in his pragma-dialectic theory – to name only a few well known works in the field. The formal logic and structure of an argument, the formalization of natural language arguments, as well as the social dynamics during communication – like the persuasiveness – have been studied extensively from different perspectives.

Research on well-designed visual support for analyzing argumentative and deliberative text has received less attention so far. Some techniques have been developed to depict individual argumentation structures to capture the formal logic or to formalize a natural language argument [22, 135, 143]. Other works cover different aspects of the communication dynamics during multi-party discourse [27, 28, 60]. However, all of these works focus on an individual argumentation or conversation. However, a social scientist cannot deduce a theory or prove a hypothesis based on an individual observation. To support the analysis of multiple arguments and deliberative processes simultaneously, suitable abstractions and aggregations are necessary. Researchers should be able to identify relationships between the functional parts of an argument and the text content, compare argumentation structures, generalize from individual argumentation structures to groups of similar structures, and analyze the deliberative quality along thread structures of written discourse.

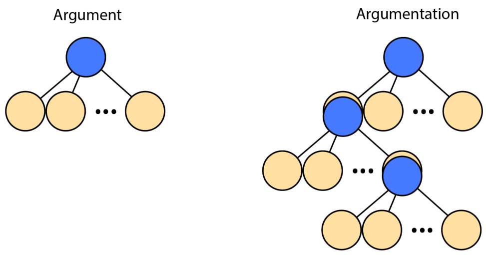


Figure 1.1: An individual argument and an argumentation consisting of multiple arguments. The circles represent argument units: blue for claims and orange for premises.

In this work, new visualization and interaction techniques are designed, implemented, and evaluated to support social scientists in the analysis of arguments with different foci: (1) the hierarchical argumentation structures with their functional parts and their order in written texts, (2) the deliberative process based on arguments in the hierarchically structured threads of multi-party online debates, and (3) the statistical evaluation of relations between the functional parts of the argument or/and its content.

1.1 Excursion into Argumentation Theory

An **argumentation** consists of a complex structure of statements that give reason to accept or reject other statements [148]. According to van Emeeren [142], "[a]rgumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge."

Argumentation models facilitate the understanding, discussion, and analysis of arguments [36, 43, 111, 140]. The argumentation model used throughout the thesis is based on Bayer [79]. The 'controversial standpoint' and the 'propositions' are statements and form the set of **argument units** for an argumentation. The 'controversial standpoint' is called **claim**; 'Propositions' are called **premises** and give 'reason to support [or reject] a claim that is subject to doubt' [150]. An **argument** consists of one claim and a set of premises that support or attack the claim. A premise of one argument can also be the claim in another argument. The set of all connected arguments is called **argumentation**. Figure 1.1 shows the construction of the argumentation structure and illustrates its hierarchical tree-like nature.

A more detailed alternative was developed by Toulmin [140] and captures real-world argumentation in which the usual rules of formal logic that

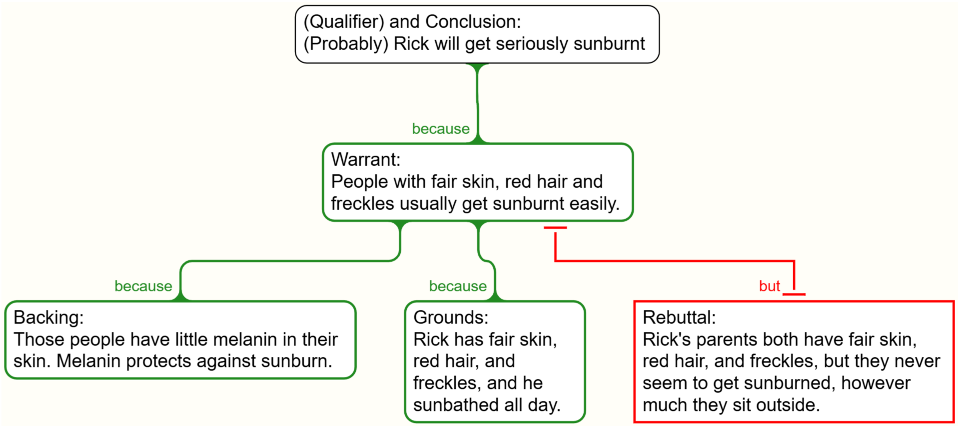


Figure 1.2: A sample argument following the argumentation scheme of Toulmin [140] modelled with the online argument mapping tool Rationale [143]. Color is used to distinguish between supports (green outline) and attacks (red outline), while the claim is shown in a blue outlined box. The keywords 'because' and 'but' help the user construct a cohesively connected argument.

had been the standard before Toulmin proposed his method may not apply. The model consists of six components:

- **Claim** The central statement or conclusion of the argument.
- **Grounds** The facts, data, or evidence that support the claim.
- **Warrant** The reasoning why the grounds support the claim.
- **Backing** Additional support for the warrant.
- **Qualifier** Indicates the strength or scope of the claim, like *probably*, *usually*, or *in certain cases*.
- **Rebuttal** Addresses counterarguments or exceptions.

Figure 1.2 shows an example argument following this model. The model establishes a framework for thinking about and analyzing argumentation in general and helps construct new arguments. But, the high level of detail makes consistent annotation difficult, as statements can sometimes not be assigned to any of the six components unambiguously. For this reason, simplified models are used in practice for computer-aided analyses of larger amounts of data. In the argumentation model of the dissertation, grounds, warrant, backing, and rebuttal are referred to as premises.

Figure 1.2 also shows how well the formalization of argumentation through argumentation models lends itself to visualization. The depiction of the functional parts of an individual argumentation together with their relations is subject to a whole branch of visualization called **Argument Mapping**. Node-link diagrams are the dominant type of visualization in Argument Mapping, since they support all types of argument models and allow translating the different argument units and relation types to respective visual elements. Argument maps have been designed for several different tasks including the visual summary of ongoing discussions (*gIBIS* [22], *ArgVis*[69], *Belvedere*[135], *DebateGraph*[91], and *Dicode*[141]), the support in decision making / contrasting points during a discussion (*Dialectic Map* [103], *SEAS* [90], and *VUE* [5]), and the presentation and analysis of a formalized argumentation (*Araucaria* [116], *ArgueApply*[112], *Argunet* [18], *Rationale* [143], and *Truthmapping* [107]). The dissertation, especially chapter 2, extends the techniques for argument analysis to the analysis of many argumentative texts at once.

1.2 Strategies for Developing Visual Analysis Techniques for Argumentation

In the process of developing visualization and interaction techniques for argumentation and deliberation analysis, new ways to systematize these techniques emerged: by the granularity of the data used for visualization and by the general task.

1.2.1 Levels of Granularity

Argumentation and deliberation can be analyzed on different levels of granularity (see Figure 1.3). On the lowest level, the **intra-text level**, the visual analysis focuses on the argumentation of one text. Tasks include the comparison of single arguments or the analysis of the order, relations, and the evolution of meta information, like topics, sentiment, or human values, in an individual argumentation. Many traditional argument mapping techniques, like *ArgVis*[69], *Rationale* [143], or *Dicode*[141], are intra-text level techniques. The thesis developed intra-text level techniques as detail views for individual essays in subsection 2.4.3 and individual threads in subsection 3.5.5.

On the inter-text, intra-corpus level, multiple argumentative texts are compared with each other. The analysis highlights commonalities and differences in both structure and metadata. A key task is the statistical evaluation of how argument units are distributed and how they (co)relate to metadata – such as stance, topics, or human values – across the entire corpus. Comparing hierarchical argumentation structures poses a particularly

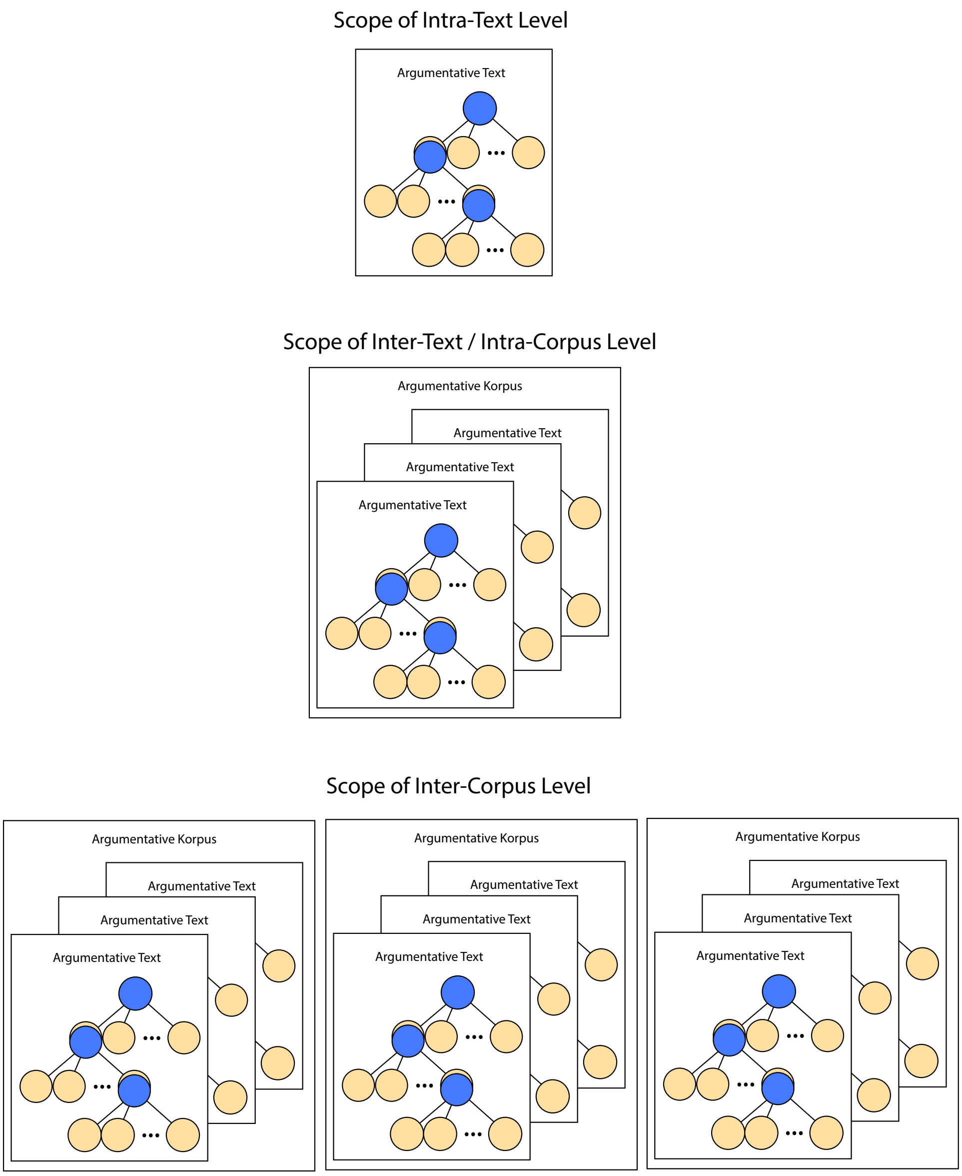


Figure 1.3: The scopes of the different levels of granularity, a visual analysis system for argumentative texts could be designed for.

interesting but still rarely addressed challenge: how to compare many hierarchies at once? This thesis addresses this question directly and provides new approaches for the mentioned analysis tasks.

The highest level is the **inter-corpus level**. These techniques compare whole argumentation and deliberation corpora and analyze how they differ in terms of argumentation structure, argumentation scheme, the proportion of argumentative text, the number of texts, or the quality and source of the contained arguments. They may support the user in their decision as to which corpus is best suited for their use case and may even allow curating datasets with texts from multiple corpora. The dissertation developed inter-corpus level visualizations to compare the user discourse of different online news platforms with each other in chapter 3.

1.2.2 Visualization Tasks

Most tasks involved in analyzing a text corpus can be categorized into two types: exploration and search. **Search tasks** include testing hypotheses and locating argumentation structures or argument units with particular properties or relations. Most standard statistics tools used by social scientists are specially designed to test hypotheses. The visual analysis techniques developed in this thesis additionally support the analyst in the interactive refinement of search queries through direct interaction with the graphical elements that represent the desired properties of the argument units, argumentation structures, comments, or threads. This step-by-step, user-driven refinement allows analysts to more easily understand the impact of each filter step and ultimately increases the quality of the final result set.

The main task supported by the visualizations developed in the dissertation is to support visual **exploration** of argumentative texts. This exploratory approach expands the analysis beyond anticipated or known information, providing a holistic view of the corpus. Inspired by the concept of the information flâneur [25], users can let their curiosity guide them through the data, following unexpected visual cues to discover new insights about the relationships within or between argumentation structures, deliberative quality, or content.

1.2.3 Abstracting Argumentation

In order to visualize an argument, a well-fitting abstraction from the raw text needs to be found. The argumentation models described in the last section define different argument units, their relations to each other, and their function within the argumentation. Applying an argumentation model to an argumentation delivers a basic abstraction that captures the **structural properties** of the argumentation. The thus captured interplay between argument unit types, stances, and the width and depth of the hierarchical structures makes for very interesting analyzes, but gains even more by considering additional information.

One often overlooked property is the **order of the arguments**. Reordering the arguments influences how an argumentation is received, e.g., whether it is coherent or persuasive [115, 122]. Most traditional visualizations of individual arguments like *gIBIS* [22], *Belvedere* [135] or *Rationale* [143] do not show the order; they focus on the logic of the argumentation. However, since in real world argumentation the order of presentation is a relevant factor for analysis both of arguments and comments, the thesis developed new visual representatives that preserve the order of arguments or comments within a structure, so that users can perceive and study its influence (see chapter 2 and chapter 3).

With the **semantic content** of the argumentation, the text and its argumentative function can be linked. The relation between the topic of a text and its argumentation structure is of interest, since it might be that from a socially grounded, emotionally loaded topic, like child marriage, naturally other kinds of argumentation structures emerge than from a more scientific topic, like renewable energies. Similarly, certain human values an argument evokes might be typical for certain topics or argumentation structures. The dissertation will take a closer look at those questions in chapter 5.

In a collaboration with social scientists from the Friedrich-Schiller-University Jena, the **deliberative quality** played a central role in the analysis. The social scientists manually curated a set of relevant indicators to measure deliberative quality in comments to online news articles [62, 96]. With this data, the evolution of individual indicators, groups of indicators (so-called dimensions) as well as the overall deliberative quality throughout an online discussion can be traced and dependencies between indicators analyzed (see chapter 3).

1.3 Research Questions and Contributions

The thesis develops, implements, and evaluates new visualization and interaction techniques that solve common visualization problems related to the domain of argumentation and deliberation research. The techniques support social scientists in analyzing argumentative texts and deliberative threads in detail, contribute insights into the domain, and provide ways to explore the data in addition to the traditional, more search-focused data analysis. The research contributions presented in chapter 2, chapter 3, and chapter 5 were published in peer-reviewed journal articles. Chapter 4 presents follow-up research in the form of a peer-reviewed short paper that has been presented at EuroVis 2025.

1.3.1 Visual Analysis of Argumentation in Essays

Chapter 2 presents the article "Visual Analysis of Argumentation in Essays" by D. Kiesel, P. Riehmann, H. Wachsmuth, B. Stein, and B. Froehlich that was published in the journal IEEE Transactions on Visualization and Computer Graphics [74]. The major research question in this work is **"How can argumentation structures be depicted so that users can compare argument strategies between (groups of) essays?"**. Employing a corpus of 402 argumentation-annotated student essays [133], the paper develops new visualization and interaction techniques to group, aggregate, and compare the multitude of tree structures that stem from the hierarchically organized argumentation within each essay. The order of stances and the width and depth of the argumentation structures speak of different argument strategies employed by the students.

The paper yields two main contributions regarding the research question: The ArguLines, a pixel-based argumentation structure visualization, and the Argument Unit Occurrence Tree, which aggregates argumentation structures. The ArguLines represent the argument units of each essay by a sequence of small glyphs. Each glyph encodes the stance, the depth, and the relative position of an argument unit. A list of ArguLines provides an overview of the corpus and can be ordered in various ways to reveal patterns and outliers. The Argument Unit Occurrence Tree summarizes the hierarchical argumentation structures of sets of essays by stacking the individual structures in a node-aligned way. The focus on the hierarchical structures facilitates the analysis of common structural features, like the position of the claim in relation to its premises (deduction vs. induction) or typical argumentation depths across the progression of the argumentation. The Argument Unit Occurrence Tree also provides means to compare subsets of essays. With the ArguLines and the Argument Unit Occurrence Tree, the paper offers social scientists two new techniques that, in contrast to traditional argument mapping techniques, are designed for comparing many argumentative texts at once and allow the user to analyze order-dependent patterns.

An additional text view serves to evaluate conclusions drawn from the other views and comprehend the annotation process. Interaction techniques for visual filtering, (sub)set selection, and comparison enable a deep analysis of essay corpora. Our expert reviews confirmed the utility of the system and revealed detailed and previously unknown information about the argumentation in our sample corpus.

1.3.2 Comment Analyzer: A Tool for Analyzing Comment Sets and Thread Structures of News Articles

Chapter 3 presents the article "Comment Analyzer: A Tool for Analyzing Comment Sets and Thread Structures of News Articles" by D. Kiesel, P. Riehmann, I. Engelmann, H. Ramezani, and B. Froehlich that was published in the journal IEEE Transactions on Visualization and Computer Graphics [76]. The Comment Analyzer supports the analysis of deliberation-annotated comment-thread corpora. It can yield insights into the distributions and relations between comment attributes, the homogeneity of thread sets, frequent thread structures, and changes in comment qualities throughout a single but in particular of multiple threads at once. The tree structure of the threads allows the transfer of some principles used to aggregate argumentation structures to the domain of communication research. But: **"How can the tree aggregation technique be extended to cope with the multitude of attributes per node studied in communication research?"**

The research contribution to the question is the design of a compound rose-chart-like glyph that enables the scientist to analyze up to 20 deliberative indicators grouped by four dimensions in relation to the political orientation of the comment. The glyphs are adaptive to different size requirements, like as nodes of the small multiple view of aggregated comment-thread structures, as well as stand-alone as visual representatives of sets of comments. The visual querying system contributes with simple yet powerful analysis and filtering mechanics to refine the sets of comments and threads displayed for study.

Additionally, we contribute a fully visual computational notebook that allows storing and comparing thread sets, preserving lines of research, and commenting on previous findings. Since the thread set representatives have the same functionality as those of the main display, each set can be the starting point of a new line of investigation.

The system design was developed in close collaboration with communication scientists in a user-centered approach. The system proved its utility in thorough reviews with the communication scientists, by corroborating existing findings in the literature, but particularly by provoking and answering new research questions. Final reviews with five independent experts confirmed these observations and revealed the potential of the Comment Analyzer for other datasets currently being created and analyzed in the communication sciences.

The design of the radial rose-chart-like glyph generated the question, whether a radius or area encoding would yield the more accurate estimations in rose charts and similar radial layouts. The study results were published in the short paper "Rose Charts: Area or Length Encoding for Fill Level of Circle Sectors?" by D. Kiesel, P. Riehmann, and B. Froehlich and presented at EuroVis 2025. Chapter 4 presents the paper. The study revealed that a mixture of both area and length estimation is naturally employed when judging the fill level of circle sectors. For low and medium values up to a fill level of about 65% area perception dominates, while for larger fill levels, a transition to length perception can be observed.

1.3.3 Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots

In order to gain a deeper understanding of relations between argument units as functional parts and other factors like the topic or the human values evoked in the text, a statistical evaluation of clusters and correlations between these attributes is essential. Scatter plots are well-suited to depict correlations and clusters for two dimensions; Parallel coordinates can depict multi-attribute relations. For an effective analysis of the statistical measures of arguments, we need both. So: **"How can we effectively combine scatter plots and parallel coordinate plots to support the visual statistics evaluation of argumentation?"**

Chapter 5 presents the article "Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots" by D. Kiesel, P. Riehmann, and B. Froehlich, published in the journal Computer Graphics Forum [75]. The article contributes a seamless transition between parallel coordinate plots, star plots, and scatter plots. The star plot serves as a mediator visualization between parallel coordinate plots and scatter plots since it uses lines to represent data items as parallel coordinates do, and can arrange axes orthogonally as used for scatter plots. The design of the transitions also motivated a new variant of the star plot – the polycurve star plot that uses curved lines instead of straight ones – to facilitate the transitions between parallel coordinate and scatter plots. Additionally, the polycurve star plot has advantages both in terms of space utilization and the detection of clusters.

Furthermore, we contribute a geometrically motivated method to embed scatter points from a scatter plot into star plots and parallel coordinate plots to track the transition of structural information, such as clusters and correlations between the different plot types. The integration of the techniques into an interactive analysis tool for exploring multivariate data demonstrates the utility of the presented approach. A user study and detailed use cases confirm the usability of the techniques. A fine-grained analysis of the relations between the argument units, the topics, and the human values used in the essays of the argument annotated student essays corpus [133] completes the evaluation.

2
Visual Analysis of Argumentation in Essays

This chapter reports on joint work with Patrick Riehmann, Henning Wachsmuth, Benno Stein, and Bernd Fröhlich at Bauhaus-Universität Weimar. It was published in IEEE Transactions on Visualization and Computer Graphics and presented at the 2020 IEEE VIS.

This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 03PSIPT5A (project Provenance Analytics) as well as by the Thuringian Ministry of Economy, Science and Digital Society in form of a scholarship.

Abstract

This paper presents a visual analytics system for exploring, analyzing and comparing argument structures in essay corpora. We provide an overview of the corpus by a list of ArguLines which represent the argument units of each essay by a sequence of glyphs. Each glyph encodes the stance, the depth and the relative position of an argument unit. The overview can be ordered in various ways to reveal patterns and outliers. Subsets of essays can be selected and analyzed in detail using the Argument Unit Occurrence Tree which aggregates the argument structures using hierarchical histograms. This hierarchical view facilitates the estimation of statistics and trends concerning the progression of the argumentation in the essays. It also provides insights into the commonalities and differences between selected subsets. The text view is the necessary textual basis to verify conclusions from the

other views and the annotation process. Linking the views and interaction techniques for visual filtering, studying the evolution of stance within a subset of essays and scrutinizing the order of argumentative units enable a deep analysis of essay corpora. Our expert reviews confirmed the utility of the system and revealed detailed and previously unknown information about the argumentation in our sample corpus.

2.1 Introduction

Argumentation is key to human communication and writing, and subject to research not only of (digital) humanists and linguists but also of computer scientists for a long time. An argumentation consists of a complex structure of statements that give reason to accept or reject other statements [148]. The study of argumentation requires in-depth insights into various structural aspects, e.g. hierarchical, stance, or sequence relations, of not only one individual argumentative text but a whole corpus thereof. While humanist and linguist researchers are often interested in the cultural or topic-based differences within these aspects, computational linguists might use them to improve the automatic detection, extraction and analysis of arguments from argumentative texts. To be successful in those endeavors, however, all researchers need to gain a deep understanding of typical and common argumentative structures as well as their differences in a corpus.

So far, visualization research focused mostly on the depiction of individual argument structures as node link diagrams (e.g. *gIBIS* [22], *Belvedere* [135] or *Rationale* [143]). Additionally, most of these techniques lack sequence information. Little work has been done for analyzing text collections (e.g. Wachsmuth et al. [55]) and only on a basic level: Even seemingly simple tasks, like determining the fraction of texts that contain a specific argument structure, are almost infeasible using such tools.

To fill this gap and support researchers in the study of argumentation, we designed, implemented, and evaluated a visual analytics system that allows the examination and comparison of structural aspects across multiple argumentative texts. It employs two novel coordinated visualizations: Our *ArguLines* summarize each individual argument structure in a glyph-based, space-saving manner whereas the *Argument Unit Occurrence Tree* (*AUOT*) overlays the argument structures of multiple essays by using extended hierarchical histograms. Elaborate queries for corpus-wide filtering of argumentative structures are enabled by direct interaction with the visual elements.

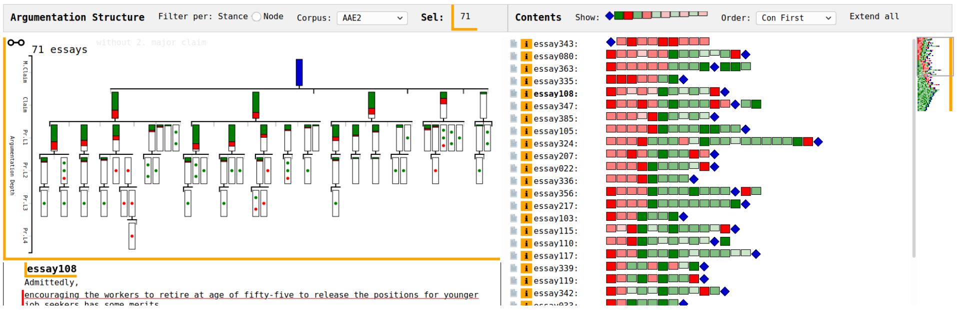


Figure 2.1: An extract from the visual analysis system for essay corpora: On the right, the ArguLines view represents each document's structure as an ArguLine in overview and detail (see section 2.4 for more details). Currently, all documents of the filtered dataset are selected (highlighted by orange markings in the list). The filter excludes all essays with more than one major claim. Essay 108 has been opened in text view on the bottom left. The structure view called AUOT (argument unit occurrence tree) on the top left provides an aggregation of all essay structures currently selected for further analysis of argumentation patterns. Each node of the AUOT shows the fraction of essays with a major claim at the node's position in blue or the fraction of pro and con claims or premises in green and red.
It is clearly visible that all essays in the selection have at least two claims (with con fractions between a quarter and a fifth), most of them three, less than half four and very few even five claims. Below that, partially filled nodes are displayed indicating more diverse structures on the premises level.

The development followed a user-centered design process that involved experts in all stages.

Our work is motivated by discussions and interviews with digital humanists and computer linguists who were dissatisfied with their possibilities to analyze text corpora with respect to argumentation practices. Digital humanists explained to us that their research includes questions such as what are the differences between the dominant argument structures of two sets of texts or which fraction of texts contains a certain argument structure. Computational linguists, on the other hand, wanted to detain the part of the structure that tends to be the same in all texts – e.g. to support finding argument structures in an unseen text – and to compare automatic structure annotations with those of human curators to assess the quality of the classification. Digital humanists also noted that it is helpful in the classroom to show students examples of student essays and good as well as debatable arguments in a visual way.

To support these and other tasks, our system for the analysis of argumentative text collections – more specifically essays – contributes the following novel visualization and interaction techniques:

- The ArguLines which represent each text by a sequence of glyphs encoding its argument units, the unit's argument depth, stance and order of appearance (Figure 2.1, right). The compact design implemented as a variant of small multiples allows the user to quickly compare individual structures with each other and assess typical patterns as well as anomalies.

- Comparing structures in an entire corpus is enabled by the Argument Unit Occurrence Tree (AUOT) (Figure 2.1, top left) which aggregates the argument structures of multiple essays into hierarchical histograms. It supports the user in determining the proportions of text that contain specific argument structures and makes it possible to examine the differences between sets of structures.

- The implemented interactive visual filtering system allows users to filter the text corpus for texts containing one or more relevant argument structures by interacting with the respective graphical representatives in a visualization.

Our reviews with external experts confirmed the utility of the individual visualizations and their coordinated interactions. This assessment was impressively substantiated by various new findings on the corpus, which were previously unknown even to experts who have conducted studies on machine learning with it.

2.2 Argumentation Model and Corpora

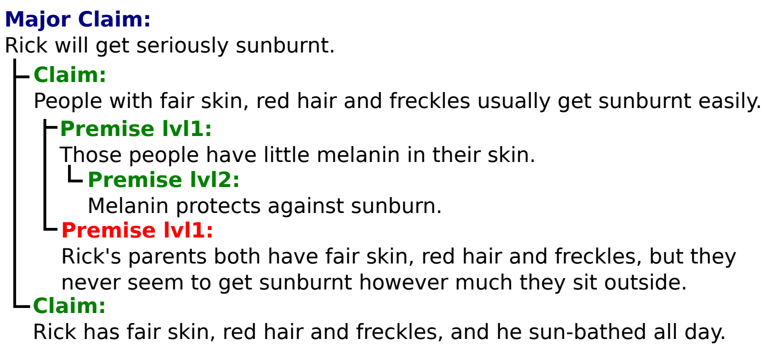


Figure 2.2: A sample argument following the argumentation scheme used in this paper. The black lines symbolize the hierarchical relations between the argument units; The labels follow the overall color scheme (Sec. 2.4): green ⇒ pro unit, red ⇒ con unit, blue ⇒ major claim.

In this design study, we extend the simple argument model introduced in the introduction and follow the nomenclature of Stab and Gurevych[133] to facilitate the communication with linguistic researchers. The root units are called *major claims* and define both the topic and stance of the essay. Their children are called *claims* and serve as starting points for single arguments arguing for or against a major claim; Further units usually contain examples, evidence or facts in favor or against their parent and are called *premises*. Since premises are allowed to relate to other premises, the depth level of a premise needs to be indicated by a level suffix if an unambiguous identification is required in the analysis (e.g. premise lvl2) (see Fig. 2.2). In simple arguments, the major claim is stated only once. In essays, however, it is not uncommon to repeat or rephrase the major claim during the conclusion such that a hierarchy may possess multiple root nodes that are each connected to all claims of the text.

Only few medium-sized, argument-annotated corpora exist: Microtexts [108] (112 texts), Argument Annotated User-Generated Web Discourse [47] (340 texts), or AraucariaDB [116] (662 argument maps). The corpus we focus on is the Argument Annotated Essays 2 (AAE2) corpus [133] of 402 essays written by English students and their prompts from *essayforum.com*. The texts are manually annotated with argument units (major claim, claim or premise), claims' stances (pro, con) and relations between claims and premises (attacking, supporting). We chose the AAE2 since it is the richest with respect to argument structure information. Additionally, we could acquire the original annotations for the two steps of the corpus creation process: In the *text annotation step*, argument units are extracted and classified as major claim, claim or premise. If multiple annotators are involved, the annotated text spans are then unified to one common consensus. In the *association step*, the stances and pro/con relationships are added to the units. These again are unified to become part of the final corpus.

2.3 Related Work

Argument structure visualization is a special form of text visualization (see Kucher and Kerren [66] for a survey). Previous work on visualizing argument structures specialized in visualizing, analyzing and synthesizing single argument structures. Most of these run under the name of *Argument Mapping* and utilize node-link diagrams to depict the argumentation: *gIBIS* [22], *ArgVis*[69], *Belvedere*[135], *DebateGraph*[91], and *Dicode*[141] serve as mind maps to support an ongoing discussion. *Araucaria* [116], *ArgueApply*[112], *Argunet* [18], *Rationale* [143], and *Truthmapping* [107] allow for analysis and evaluation of an argument. *Dialectic Map* [103], *SEAS* [90], and *VUE* [5] facilitate the decision making process during a discussion by automatic analysis and visualization. While effective for the tasks they have been created for, they do not support comparisons between multiple texts or consider the order of units in the argument. Wachsmuth et al. [55] provided a static, accumulated argument structure visualization for comparing sets of texts. However, it is very hard to read for the non-expert user and lacks clarity due to clutter and overplotting. Thus, we provide interactive and new simplified visual representations for accumulated argument graphs and to compare the structures of argumentative texts.

The problem of visualizing multiple argument structures can be interpreted as the problem of visualizing multiple trees. In that area, the *DAViewer* [162] facilitates the comparison of the discourse trees created by different discourse parsers using interactive dendrograms. Similarly, Bremm et al. [12] and Munzner et al. [100] find and display similarities and differences in phylogenetic trees. Instead of using small multiples of the full tree like the *DAViewer* and the works of Bremm et al. [12] and Munzner et al. [100], we introduce small multiples of glyph-based tree summarizations accompanied by an accumulated view for detailed comparisons, since Graham and Kennedy [45], who surveyed different strategies of tree visualizations, concluded that agglomeration is the most suitable for comparative tasks. Viewed from another angle, parts of an argument structure can also be seen as a sequence of events. Monroe et al. [99] developed a system for sequential pattern analysis in medical records and Cappers et al. [15] explored the possibilities of visualizing event sequences containing multivariate data. Riehmann et al. [121] show the sequence of plagiarism events in a text as DiffLine. All approaches are close to the ArguLine overview visualization, but – like the ArguLines – are not able to represent the hierarchical nature of argument structures in detail. Therefore, our proposed system implements a coordinated multiview approach to present all aspects of the data without overloading a single visualization.

Wachsmuth et al. [52, 53, 54] developed a visualization of the sentiment flow in hotel reviews. Similar to the argument flow in the ArguLines and structure view, it shows the changes in sentiment throughout an argumentation, but fails to present them in a way that allows the comparison of a whole corpus of texts.

Spoken argumentation is analyzed by the VisArgue group of the Konstanz University. They specialize in visualizing transcribed discussions, showing the topical development and deliberative quality of a discussion over time [28, 42], finding patterns in a discussion [64], or summarizing a discussion in a mind map [28]. Outside this group, *Conversation Clusters* [7] summarizes discussions in topical word clouds, and *GroupMeter* [88] supports students to develop argumentation skills. Group discussions, however, often contain multiple intertwined threads with separate argument structures. To develop a first argument structure analysis system, we concentrated on the single-threaded, monological argumentation in essays.

Stance visualization has been studied extensively on other kinds of texts, foremost social media. StanceXplore [94] allows the study of stances in social media over multiple attributes, like time, space, topic and content similarity. *MultiConVis* [60] shows the timeline and stance distribution of discussions in internet forums. DoSVis [84] enables research on stance in longer documents. An extensive overview is given by Kucher et al. [85]. Unlike current stance visualizations that focus on the demography of writers, topics or temporal development, our system allows the study of stance and its role within arguments.

2.4 Argument Structure Analysis System

Our system for the visual analysis of argument structures is designed to support (computational) linguists and digital humanists in analyzing the argument structure of a corpus of essays. We started our development by deriving an initial list of required functionalities for our system in initial interviews with domain experts about their usual tasks:

Task 1 The task our experts conduct mostly is *finding common structural and temporal patterns* within a given corpus, since it is the most crucial part in studying argumentation. It is also highly relevant in building better heuristics for automatic structure detection. Students might benefit by developing more sophisticated argumentation skills.

Task 2 The second most important task in argumentation studies is *comparing the structures of sets of text* with each other for revealing e.g. the differences between the typical argument structures of authors with different cultural backgrounds. Students might also benefit from visually exploring the differences of well written and debatable essays.

Task 3 Support for *finding texts containing specific or potentially uncommon structures* is crucial for verifying hypotheses about the corpus (e.g. do texts beginning with a con argument argue against the given prompt?)

or comparing defined subsets of essays (e.g. all essays with one major claim against all with two or three major claims).

Task 4 To generate meaningful results, the quality of the corpus is of utmost importance. Therefore, means to *detect outliers* and *compare the original annotations* are needed.

Task 5 A system supporting scholarly work needs a way to *read/extract suitable text passages for evaluation or as prove by example.*

Currently none of these tasks are easily manageable using regular node-link diagrams. Especially comparing argument structures is difficult since the layout of the diagrams is not unified and the node sizes depend on the length of the embedded argument unit's text. So, even similar structures tend to look different in traditional tools.

To overcome this deficiency, we developed a system consisting of three interactively coordinated views (see Fig. 2.1): (1) The ArguLines view on the right summarizes the entire corpus in a Overview and Detail list of ArguLines which represent an essay as a sequence of glyphs, one for each argument unit. The list can be resorted to *find common temporal patterns* or *detect outliers*. Additionally, it provides detail on individually selected essays on demand, e.g. on their *original annotations*. (2) The structure view on the top left allows the analysis of the argument structures of the texts selected in the overview list using the hierarchical histograms of the Argument Unit Occurrence Tree (AUOT). It is used to *find the most common structural patterns* and to *compare the structures of sets of text*. (3) The document view on the bottom left provides means to *read/extract suitable text passages* connected to the abstractions of the other views. Each visualization can be used as a filter to reduce the corpus to *texts containing specific structures*. The ArguLines view is the pivotal visualization in our system since it is used in common tasks which first scrutinize the list of essays for potentially relevant texts and select subsets for closer inspection in the structure and document view. Therefore, the layout was chosen such that the ArguLines have enough space to be used to maximum effect. The ArguLines exploit the whole height of the screen and, since the ArguLines' overview also functions as a scroll bar, is placed on the right side of the screen. The other two views have been arranged such that the structure view typically used in an earlier step of the analysis is on top.

All visualizations follow a system-wide color design to differentiate between different argument units as well as selected subsets. The color scheme for the argument units follows the European standard: green for pro units,

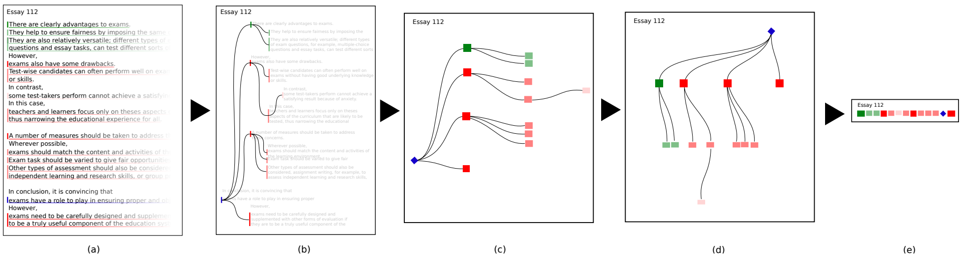


Figure 2.3: Construction of an ArguLine. (a) The annotation of a given text contains the identified text chunks as well as their stance and type (pro or con; major claim, claim, or premise). (b) The relations between the text chunks specify a hierarchical argument structure. (c) The text chunks are transferred into glyphs that encode the stance (pro ■ or con ■) and the type of argument unit (major claim ◆, claim ■ ■, premise lvl1 ■ ■, premise lvl>1 ■ ■). (d) Following the mental text flow model of our experts and the orientation of the commonly used node-link diagrams, the graph is rotated; now depicting the text flow horizontally and the depth vertically. In the last step, the tree is vertically compressed into a line while preserving the order of the units within the text.

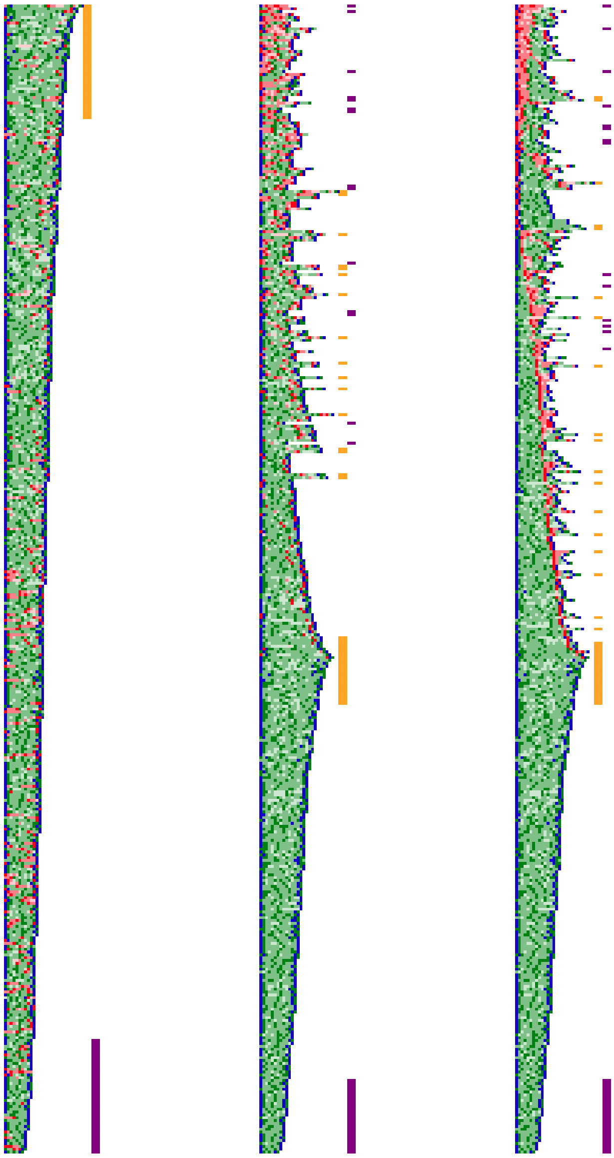
red for con units. The major claims – as the units that express the essence of the essay – are separately marked in blue. The lightness is mapped to the depth level within the hierarchical argument structure: the fainter, the deeper. This scheme was chosen as conceptually, the deeper a unit is within the hierarchy, the less influence it has on the overall argumentation. Therefore, it should be less prominent in the visualizations. Unfortunately, this pro/con model is not compatible with the most common form of color blindness (red/green). To cover these cases, the color scheme is customizable via a configuration file. The colors to denote selected subsets were chosen such that they are easily distinguishable among themselves and do not intersect with the set of colors for the argument units.

2.4.1 Overview: Corpus Summarization with ArguLines

The overview is conceived as list of all essays in a corpus that gives an impression of its scope and of the frequency of certain temporal patterns in the argument structures it contains. Detail is available on demand. Each essay is represented by an ArguLine which abstracts the argument units from the concrete text and compresses the hierarchical structures into a single line which provides a compact impression of the text flow and depth within the argument structure.

An *ArguLine* is constructed by replacing each argument unit with a representative that encodes both its stance and depth level (see Figure 2.3). Similar to common argument maps, claims and premises translate into rectangles, but with a height related to the depth level and a color that encodes both the stance of the unit in hue (pro $\Rightarrow$ green, con $\Rightarrow$ red) and the depth level in lightness (the deeper, the lighter the color). Major claims are treated specially since they compose the most important units in an essay defining the topic and stance of the whole text. In many essays, they figure more than once, mostly within the introduction and/or conclusion. To have them pop out among other units, major claims are redundantly encoded as blue diamonds instead of rectangles. This way, the number and positions of major claims can be retrieved at a glance just as well as the number of overall argument units, the distribution of pro and con units, and the distribution of depth levels. The visualization does not show non-argumentative text pieces, since they do not convey any structurally relevant information. If needed, the information can be shown on demand using the length-proportional detail view of the ArguLine.

The ArguLines reveal overall tendencies, like the blue major claim at the start and end of most essays (see Figure 2.4), and subsets of texts with similar structure: Some texts do not contain any con units, some scatter them, and some present them as one block. From these subsets and tendencies, samples of interest can easily be identified and selected for detailed examination either with the other visualizations (subsection 2.4.2, subsection 2.4.3) or the ArguLine's own detail views: A tree view to provide the full argument structure of a single text, a length-proportional view to analyze the actual positional distribution and lengths of the argument units, as well as an annotation view to review the steps and decisions made during corpus creation if the data is available. The tree view (Fig. 2.5, b) is brought forward on clicking the ArguLine's info icon. On activation, the ArguLine's glyphs move vertically to their respective depths within the argument hierarchy. Additionally, lines appear to explicitly define the structural relations between the argument units. The explicit argument tree strengthens the mental model of the analyst and resolves uncertainties introduced through the reduction of the full structure into an ArguLine. The length-proportional view (Figure 2.5, c, top line) is activated similarly for a single ArguLine or via a control element for the whole corpus at once. The glyphs move to their relative positions in the text and stretch or compress to match the length of the text chunk. Additionally, gray rectangles are included to show non-argumentative text. The positions and lengths are encoded relative to the length of the longest essay showing the length-proportional view, enabling comparisons between texts. Including the unit's length and non-argumentative text revealed that there is often a non-argumentative introductory text and hardly any non-argumentative passages between argu-



(a) (b) (c)

Figure 2.4: Different orderings of the list of minimized ArguLines in a typical three-step analysis of the AAE2: (a) ordered by the number of argument units. In (a), the selection of the 40 longest (orange) and 40 shortest (purple) texts was conducted. (b) ordering by the number of con arguments in the text shows long essays (orange) are more seldom purely pro than short ones (purple). (c) ordered by the position of the con units shows short essays (purple) place con units earlier in the text.

ment units. It also showed that premises are mostly longer than claims or major claims. The annotation view (Figure 2.5, c) shows the originally an-

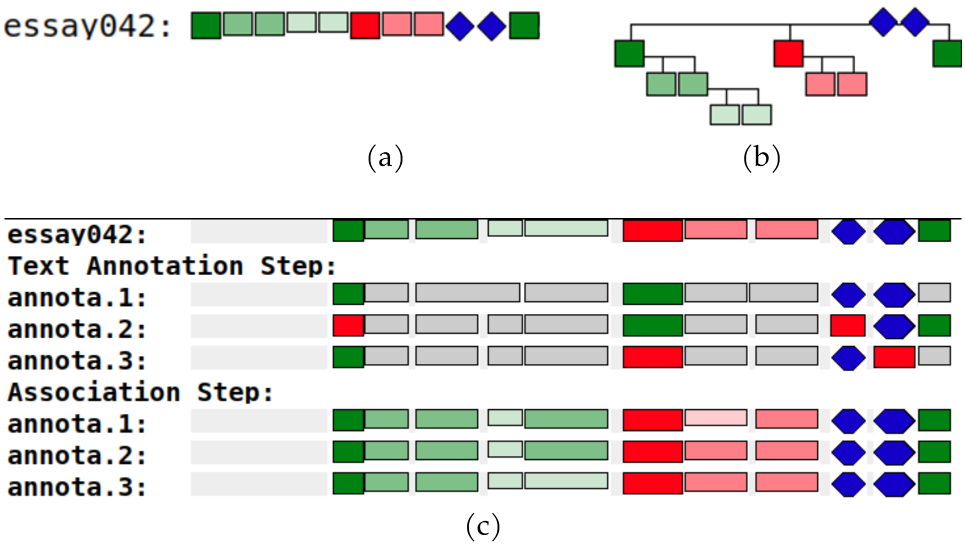


Figure 2.5: The (a) regular ArguLine for essay 042, its (b) tree view and its (c) annotation view (including the length-proportional view in the first line) showing segment lengths, text positions as well as the choices of all annotators when marking up the argument structure during the text annotation step and the association step. During the text annotation step, annotator 2 and 3 are in consensus about the positions and lengths of the argument units, but never agree in type. The color code follows the overall color scheme, adding gray for units that do not have a stance assigned yet and light gray for non-argumentative text.

notated text chunks of all annotators – in case of the AAE2 corpus: three – in both annotation steps if that data is available. The visual encoding is equivalent to the encoding in the length-proportional view. The annotations of each annotator are shown as one line for each step. The annotation view provides valuable information about the consensus process of the annotators; annotations in a text with many disagreements might be less reliable. The example in Figure 2.5 shows that annotators 2 and 3 disagreed on the type of most units while agreeing perfectly on their boundaries during the text annotation step. In the association step, the relations are mostly agreed upon; only two disagreements concerning the depth of the fifth and seventh argument unit are visible.

Reordering the list allows for visual clustering of structurally similar texts and the detection of patterns. Fig. 2.4 shows a typical three-step analysis of the whole AAE2 corpus: in (a) the number of argument units is key to the ordering. In this view also, the selection of the 40 longest (orange selection) and 40 shortest (purple selection) texts has been conducted. The corpus can be reordered by the number of con arguments in the text (b). More than half of the short essays (purple) do not contain any con units, while the fraction

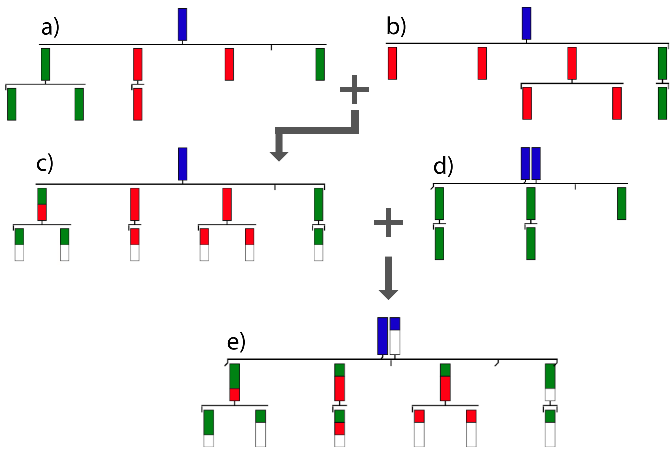


Figure 2.6: Construction of the Argument Unit Occurrence Tree (AOUT). The aggregation of the two argument structures a and b by merging positionally equivalent nodes that render the number of overlapping pro and con units at the respective position as one-bar histogram (c). Adding d extends the aggregation resulting in an AOUT (e) surveying the argument structures of 3 essays.

of purely pro essays is much smaller in long texts (orange). Interestingly, the texts with the most con units are, nevertheless, within the shortest texts (purple). When reordering by the position of con units within the argumentation(c), the list shows con units show up earlier in short essays (purple).

2.4.2 Structure View: Aggregation with AUOT

Argument structure views support the researcher in detecting common structural and sequential patterns, analyzing texts containing specific substructures, and comparing subsets of essays against each other. Until now, researchers had to laboriously compare the node-link diagrams from each relevant argument structure with one another to accomplish any of these tasks. So, a view that can aggregate a subset of argument structures across multiple essays would simplify this task considerably.

Early efforts followed the initial suggestion of one of our internal experts to overlay the logical graph structures of all essays visually into a single aggregated graph visualization while maintaining the full sequential position information of each argument unit (not depicted but in principle an improved version of the visualization shown in Fig. 2.12). However, it led to misinterpretations and did not allow deriving reasonable conclusions for the structural analysis (e.g. for ascertaining the characteristics of a set of texts such as the typical number of major claims/claims/premises, the changes in stance between subsequent claims, etc.). Thus, we developed

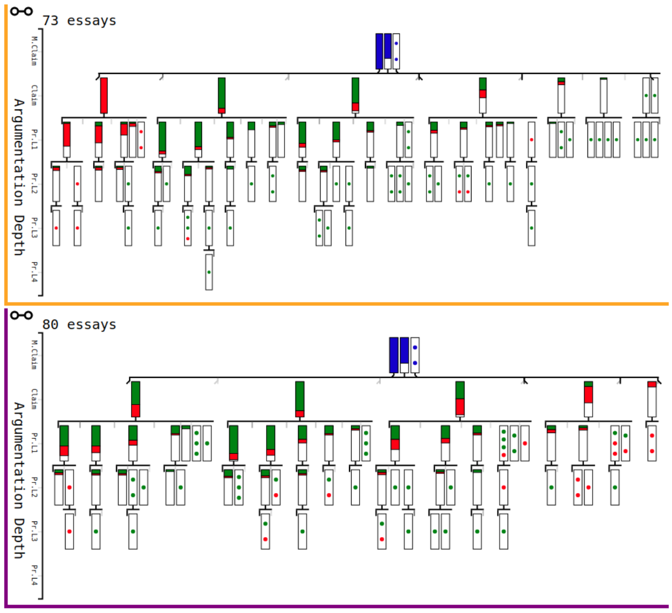
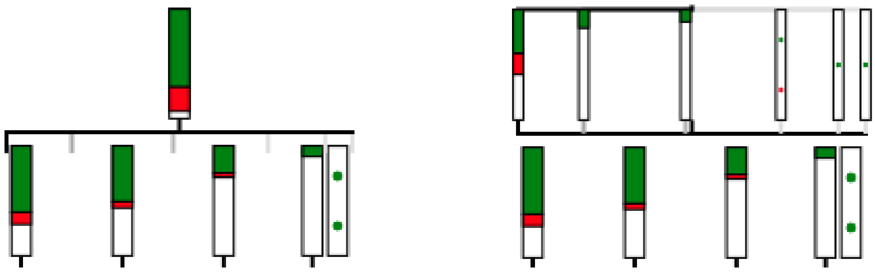


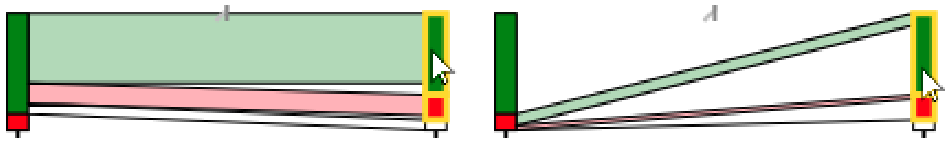
Figure 2.7: Comparing 73 essays beginning with con arguments (top) with 80 essays ending with con arguments (bottom) using the AUOT visualization. All essays have at least one major claim (the first node at the top level is fully filled with a blue bar), about three quarters of the essays have a second major claim and two essays have a third one (there are two blue circles in the third node). In the upper AUOT, there is one essay that has 7 claims, while in the lower AUOT, essays have only up to 5 claims.

the *Argument Unit Occurrence Tree* (*AUOT*) which uses the new concept of hierarchical histograms and focuses on the hierarchical organization of the aggregated argument structures (see Figure 2.6). The histogram of each node in the AUOT, displayed as a rectangular box, indicates the fraction of essays within the selected set of essays that have a node at that respective structural position. The colors of the bars represent the fraction of pro (green) and con (red) units while the white part represents the essays that do not have a node at that particular position. If a bar gets too small to be readable, it is replaced by likewise colored circles, which each represent a text of the corpus and are called countables in the remainder of the paper. To emphasize the hierarchical structure, siblings are connected by a dark horizontal line. Small vertical stubs indicate the location of the parent node in relation to the child nodes and the gray level indicates the probability of that location. Figure 2.7 shows two subsets of essays as AUOTs that were previously selected in the ArguLines overview and can now be used to compare a set's structural characteristics. The most prominent difference between the sets is the pro/con ratios of the claims. In the upper,



(a) (b)

Figure 2.8: Detail view of the third claim and its children in Fig. 2.7 (top) (a) showing the normal node and (b) analyzing the probabilities of its positions by replacing it with multiple, slimmer subnodes for each position: the claim occurs most often before all its children but may occur at any other position as well (with far smaller probability).



(a) (b)

Figure 2.9: Detail view of the second and third claim in Fig. 2.7 (top) analyzing the evolution in stance between the two nodes with cross-count ribbons: (a) about a fifth of the pro claims turn into con claims and (b) more than two thirds of the con claims turn into pro claims. The proportion of second claims that do not have a third claim (the white band) is very small (about 5% of the selection) and stems mostly from pro claims (a).

orange subset, all essays start with a con argument (the bar of the first claim is completely filled in red), while the lower, purple subset shows a higher probability of a con claim later in the text (from the third claim on, the red bar is larger than the green one).

The stubs only hint at the parent's most common position. A more detailed analysis of a parent node's position in the texts in relation to its children can be revealed on demand. Right-clicking on a node leads to its replacement by (multiple) slimmer subnodes at the actually existing positions of the node in relation to its children (Figure 2.8).

The cross-count ribbons enable the analysis of sequential patterns of changes in stance. They can be drawn between any two nodes in the AUOT. Figure 2.9 analyzes the similarity of the stance ratio between the second and third claims in the upper selected subset. One might expect that pro units would be followed by pro units in the next claim, and con units would be followed by con units. Spanning the cross-count ribbons between the pro

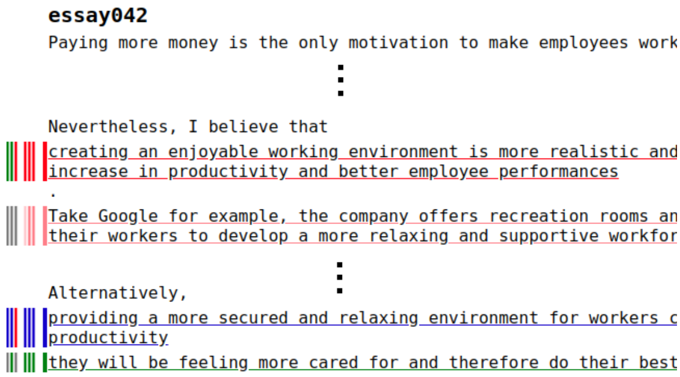


Figure 2.10: The document view of essay042 with some text cut off at the right side. Within the text, argument units are underlined. On the left, vertical bars show the markup of the three annotators in the two annotation steps and the resulting consensus. In this example, disagreement among annotators occurred with respect to stance and depth of argument units.

part of the second claim and the third claim shows that actually, about one fifth of the pro units become con units and a small fraction does not have a third claim at all. Both subnodes and cross-count ribbons have been implemented to only appear on demand since they convey supplementary information that would clutter the visualization otherwise.

2.4.3 Document View: Text Evaluation

The document view (Figure 2.10) allows the user to check for the validity of found patterns in the other views, pick examples for different subsets of essays in the corpus, and verify the annotations of the argument units. It shows the full text of one or more essays on demand at the bottom left of the screen (Figure 2.1). Within the text, all argument units are highlighted with an underline and a vertical bar on the left hand side. The color scheme matches the colors used with the ArguLines. The color of the bars encodes the stance of the unit towards the overall stance of the essay (major claim ⇒ blue, pro ⇒ green, con ⇒ red) as well as the hierarchy level (the lighter the deeper in the hierarchy). The visualization can be extended to reveal the original annotations, as well (Figure 2.10). Besides the vertical bar showing the agreed stance and depth of an argument unit, vertical bars reveal the annotators' markings during both steps of the annotation process. The leftmost cluster of lines shows the choices of the text annotation step and the second cluster shows the annotations of the association step. The text lines are wrapped such that the information provided by a vertical bar on the left applies to the entire corresponding text line. Important characteristics of the original text such as the placement of paragraphs and text flow are still preserved. To lead the user's attention to the interesting annotation cases, we do not show annotation data in sections in which all annotators concurred with each other.

2.4.4 Visual Queries, Filtering and Linking

One idea of our experts was to shape queries visually to find similar essays corpus-wide. Their discussions mainly revolved around two types of queries: queries regarding the argument's hierarchical structure (e.g. "I want to see all essays of a certain complexity/depth and width", "Show me all essays with at least 2 major claims") and queries regarding the text flow (e.g. "I need all texts starting with a con claim"). Our visualizations are ideally suited for these types of visual queries since they exactly match our two types of visualizations. The AUOT expresses aggregated hierarchical structure, hence, its elements can be used to link and filter essays based on type, stance and structural position. The ArguLines show sequential data, thus allowing for type-, stance-, and sequential position-based queries. An ArguLine expanded to tree view features a mixture of both. As a result, we implemented a consistent set of filtering and linking techniques that acknowledges the different types of encoded information (hierarchical structure and text flow) in our coordinated multiview visualization.

Hovering over an argument unit in the ArguLine view, for example, highlights all matching units at the same sequential position in the text, having the same type (and optionally also the same stance) in all other ArguLines. At the same time, the AUOT and text view highlight the positions of the same argument units in the hierarchical structure and opened texts respectively. On the other side, hovering directly over a particular node (or optionally the pro or con stance inside this node) in the AUOT reveals all argument units sharing the same path through the tree (and optionally the stance of the hovered node) in the ArguLines which highlights the hovered unit's sequential position individually in every ArguLine (Fig. 2.11) and the corresponding text parts in the text view. Both highlighting modes affect the ArguLines overview which slightly fades all non-matching units so that the number of essays containing a matching unit can be estimated at a glance. To quickly find the very ArguLine that corresponds to a text or connect the text of an argument unit to its structural position, hovering over any of text units in the text view will highlight that particular unit in its ArguLine as well as in the AUOT.

The highlight on hover also serves as a preview to filtering (which is invoked by clicking an argument unit's representative), generally removing all non-matching ArguLines and morphing the AUOT to display only the subset of matching essays. Multiple filters can be applied on top of each other to further narrow down the set of essays. Each filtering step can be undone any time, by clicking on the filtering argument unit's representative (marked by a thicker black outline). An inverse filter (removing all essays that contain the particular argument unit) and a selection-specific filter (filtering only within a selected subset of texts) complete the implemented analytics toolset.

The visual filtering system allows for shaping complex queries by simply clicking the structural elements that should or should not be part of the subset. Using comparable filtering and querying systems like SQL, finding documents that figure e.g. only one major claim and a con claim at first text position would require a complex SQL statement involving several GROUP BY and WHERE clauses while using the AUOT and ArguLines only 3 clicks are necessary: one on the second major claim (while holding the control key to activate the inverse filter), one to show the ArguLines tree view, and one on the con claim at the first position in the tree view.

2.5 Expert Reviews and Domain Findings

The developments took place based on a user-centered approach, in which two local experts evaluated the status quo at regular intervals and provided feedback for further development. After completion of the main functionalities, we additionally asked three external experts, who were familiar with the used corpora, to evaluate our system. Subsection 2.5.1 describes the feedback about the system itself and subsection 2.5.2 provides an overview of the most relevant linguistic findings revealed in these sessions.

2.5.1 User-centered Design and External Reviews

For the user-centered approach, we conducted regular reviews with two local experts during development (at least five times with each expert). Both are very experienced researchers working as computer scientists in computational linguistics and text retrieval with more than 15 respectively 10 years of experience.

After the dissatisfaction regarding early prototypes showing aggregated structures (not depicted but in principle improved versions of the visualization shown in Fig. 2.12) the introduction of the ArguLines was a turning

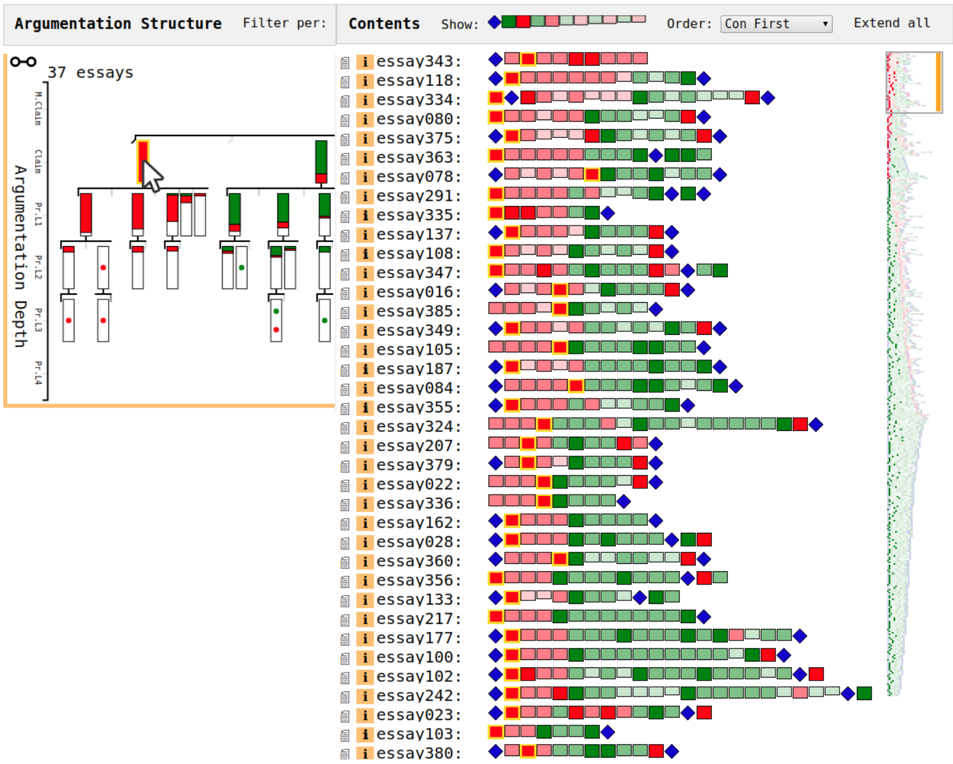


Figure 2.11: Hovering over a node in the AUOT reveals all argument units sharing the same path through the tree in the ArguLines.

point in the third round of discussions – for our experts as well as for ourselves. The experts stated that the ArguLines are far better suited for the task of getting an overview of a corpus. There was even a revision that arranged the Argulines in a Focus and Context approach, but we decided for the Overview and Detail layout in the end since it was preferred by the experts. However, the ArguLine Overview and Detail could not reveal all relations in a corpus that our experts were interested in, e.g. whether connections exist across different argument substructures. For instance, the graph visualization by Wachsmuth et al. [55] for the AAE2 corpus indicated that premises were locally clustered around their claims and the premises of an argument did not cross-reference the claim or premise of another argument. However, the legibility of this graph drawing severely suffers from clutter and overplotting. In order to investigate such relations better, we developed the AUOT visualization (Section 2.4.2).

For evaluating the final system, we conducted expert reviews with three independent experts: a computer scientist working with the group that compiled the AAE2 corpus [133], a linguist who was involved in gathering the Microtexts corpus [108] and a second very experienced linguist also familiar with both corpora. All of them are familiar with computer-assisted research and use basic visualizations in their daily work. During the reviews, the visuals of each part of the interface were explained separately and directly demonstrated with sample tasks that they are typically faced with in their research, like confirming or contradicting hypotheses they had about the corpus. The gained insights were discussed before moving on to the next feature. To a certain extent, the responses were similar. All of them were very enthusiastic about looking at all essays and their argument structures in a corpus at once using the ArguLine overview. They used the different sorting options extensively (Section 2.4.1) and were fond of the patterns and argumentation strategies (Task 1 in Section 2.4) that were revealed as a result. One expert stated: "This is really great, I can immediately see the distribution of pro and con-arguments.", one expert declared enthusiastically, "You can see very well how different people argue. Some keep themselves short, try to be succinct, and yet illuminate both sides, others like to hear themselves talk and write endlessly only about their own point of view." The corpus overview also allowed the expert to visually filter the texts based on the shown structural patterns (Task 3), hence he continued: "As a person with little time, I can immediately see which texts I would read. I would leave out the long, one-sided ones from the beginning." meaning long texts without any con arguments at all; a large fraction of the entire corpus. Another expert stated: "I would primarily use those [the ArguLines]. I have the sequence and with the icons the argumentational depth as well." One wished for being able not only to sort but also to group the ArguLine list by entering his own criteria.

The pro/con colors were obvious to all of them. Two were very fond of mapping the depth of the tree nodes as color shades and thus revealing the level of importance of an argument unit. The two linguists also immediately understood the potential of the AUOT for comparing individual subsets of the corpus against each other with respect to pro/con patterns. Here, one suggested being capable of generating subsets by providing a criterion (see sorting above) instead of selecting them individually in the list. For the computer scientist, the pixel-based overview was more revealing since he realized immediately how large the fraction of essays in the AAE2 corpus really was that did not contain a single con argument. For investigating this further, we immediately had a closer look at some of the wordings in the texts guided by the argument unit linking provided across the visualizations. All experts considered it essential to have direct and instant access to the text of the argument units.

Overall, they stated that our system would facilitate the process of deriving and verifying hypotheses. For instance, the experts told us that the usual way of checking hypotheses in linguistics involved a lot of tedious manual work. They would look at individual texts or, if available, so-called RST trees (Rhetorical Structure Theory by Mann [139]). In both cases, they would have to deal with multiple independent application windows while we provide linked views in our system. Furthermore, one of them suggested that there is a large potential in extending our ArguLine concept for providing an overview of all RST trees [104] of a corpus at once. This expert also would like to use the system for teaching linguistics and to show students the particular features of the Microtexts corpus.

Even though the five final expert reviews confirm the utility of our system, we are aware that humanists who are not used to computer-assisted research and visualization systems may need a longer training phase than our experienced researchers. However, the suggestion by one of our external experts to use the tool for teaching linguistics shows that the expert felt that it should be quickly learnable.

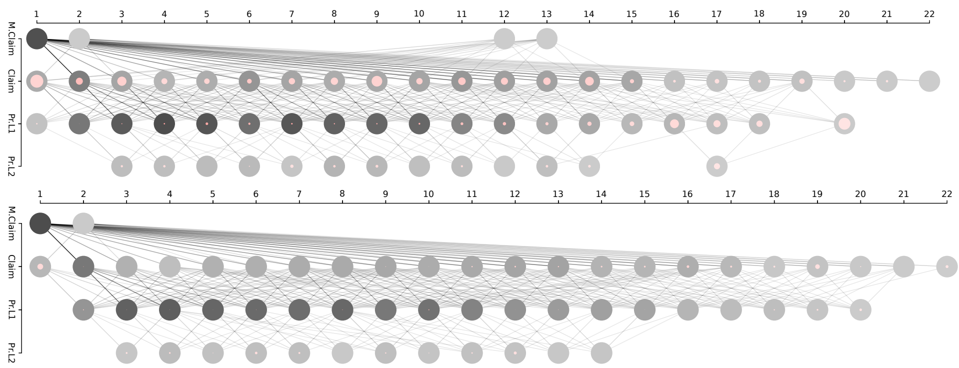


Figure 2.12: Comparing structures of texts with myside bias (bottom) with those without bias (top) with the early visualization concept used in Wachsmuth et al. [55]. In this version, argument trees are laid out in their sequential order (identical to the tree view of the ArguLine) and plotted on top of each other, aggregating units of the same position and depth. The lightness of the circles representing the nodes indicates the number of units in the aggregation; the darker the more units. Some nodes have light red inner circles whose area compared to the overall area of the node depicts the ratio of con units at this position. Nodes with very few occurrences are removed.

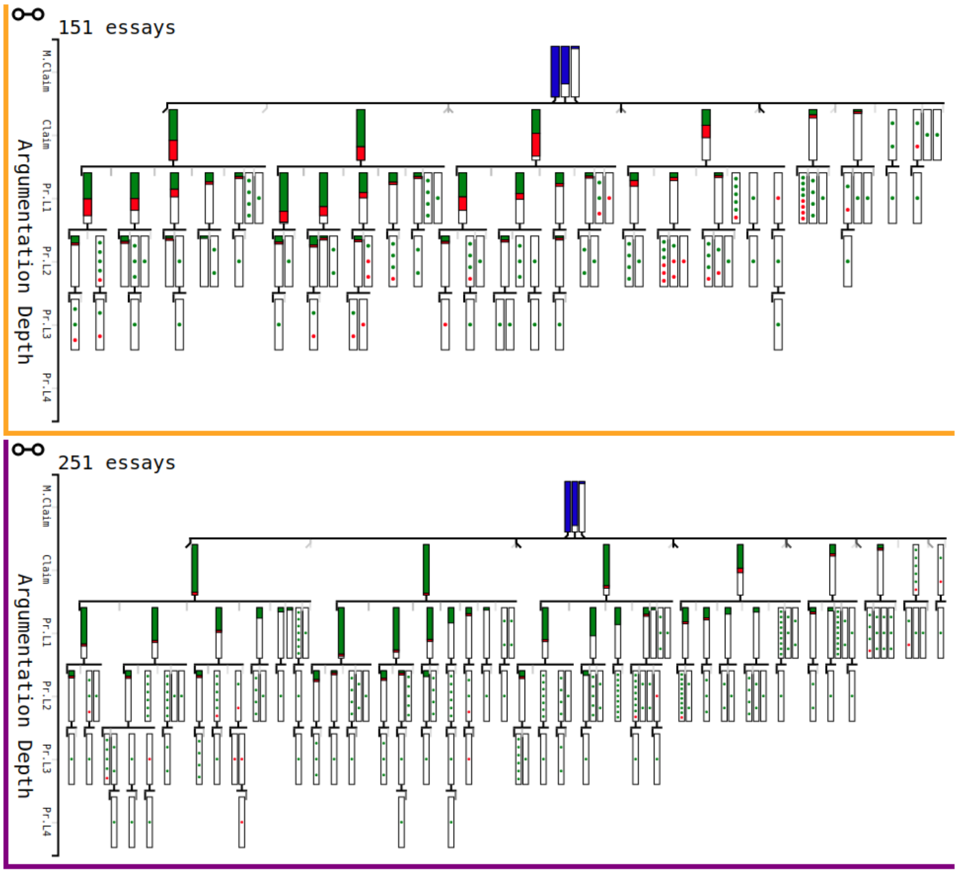


Figure 2.13: Comparing the same structures as in Fig. 2.12 (with myside bias at the bottom, without myside bias on top) with the proposed AUOT. The argument's hierarchy is far better visible, the number of each type of unit (major claims, claims, premises) countable and the frequency of each unit position interpretable. However, positional information is less present in this visualization.

2.5.2 Linguistic Findings

Using our system resulted in several intriguing insights about and into the AAE2 corpus [133] that were previously undiscovered by our experts – despite their thorough investigation and knowledge of the corpus.

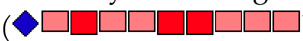
With the computer scientists who used the initial version of the ArguLines, we observed different patterns (Task 1) in the sequences and the argument structure of the essays in the corpus. A more thorough investigation using the AUOT's filtering techniques revealed e.g. that 18% of the essays have 1 major claim, 78% have 2 and 4% have 3; 57% contain at least one con unit and 43% do not; the pro/con ratio is exactly balanced for only 2.5% of the essays, whereas 92.5% lean towards pro and 5% towards con; about 75%

of the essays start with a major claim, 10% with a con argument unit and 15% start with pro. The first claim, however, is pro in 82% of the corpus, and con only in the remaining 18%. None of these subsets with their particular properties became apparent in prior analyses of the corpus.

By analyzing the pro/con ratio of essays (Task 1), together with the expert who was involved in compiling the AAE2, we found that 173 of the 402 essays contain heavy cases of myside bias[110] and do not contain any con argument units at all. Cases of less heavy myside bias can be studied using the categorizations of Stab and Gurevych [132] and Wachsmuth et al. [55], where an essay has myside bias if it does not have a con claim in its main body (meaning without the first and last paragraph). Comparing the cases of myside bias according to that categorization to all essays without myside bias (Fig. 2.13) reveals a similar general structure: both sets always have at least one major claim, at least two claims and the odds of having a second/third major claim or third/fourth/fifth/etc. claim are about the same. Identical to the original conclusions drawn from the original visualization (Fig. 2.12) by Wachsmuth et al. [55], the visualization suggests the first major claim to be usually stated before its claims in both sets. However, the AUOT allows the additional insight that an essay may contain multiple major claims and how likely it is to encounter a second and third one. Furthermore, Wachsmuth et al. [55] found that should the major claim not be in first position, the text is likely to start with a con argument; the AUOT, however, allows for statements about the actual probabilities: about 3/4 if the major claim is placed after the first claim, about 1/3 when it is after the second, and about 2/3 when after the third. Furthermore, the ArguLine list clearly shows that the con sections of most texts cluster either at the start or end of the essay in the vicinity of a major claim.

An analysis of the subsets with respect to the number of major claims (Task 1&2) was also conducted with one of the linguistics experts. It revealed that essays with only one major claim place it towards the end of the essay summarizing the point of view of the writer. The preferred position of the first major claim – it may be repeated up to two times throughout the text – across the entire corpus is the very beginning to introduce the stance of the author along with the topic of discussion. The interplay between the argument units in the ArguLines and their textual counterparts explored in the expert reviews lead to the insight that the repetition of major claims was in most cases not done by simple rephrasing as one might expect. Instead, the restatement of a major claim is typically more general, more specific or raises another but strongly related point, e.g. "[...], we cannot ignore its [meaning the introduction of machines and technology] negative effects" and "[...], the problems posed by machines should draw more attention, as well".

Investigation of the order of claim stances during the argumentation (Task 1) allowed interesting insights as well: Using the cross-count ribbons between subsequent claims, we found a surprisingly high probability for a change in stance between them (pro→con: ∼20%, con→pro: ∼80% between first and second claim, pro→con: ∼50%, con→pro: ∼50% between second and third claim, etc.). This pattern of changing stances, as one of the linguists explained, is called *thesis-antithesis pattern*. Furthermore, the last claim of an argumentation often presents a con stance. As it seems, the students liked to close with a weak con argument to further elevate the major claim.

The detailed visualization of the annotation process (Task 4) with the same linguist revealed the attention the authors of the corpus paid to annotator disagreements. With three annotators, one would expect to see a majority vote after each annotation step. Instead, in more than half of the cases, the vote of one annotator was further used or even a completely new structure was derived through discussions with all three annotators and became the agreed one. In general, we found a number of annotation inconsistencies during software development and expert reviews. One case that would have been difficult to find if it was not for the ArguLines and their sorting features (here *pro/con ratio*) was an essay containing only con arguments without any pro argument at all (). One linguistics expert decided to examine the origins of this outlier and took a closer look at the corresponding essay (Task 5). Without the need to read the entire essay, the explicit highlighting of the argument units (see Figure 2.10) in the text view and the interactive linking to their counterparts in the ArguLine was crucial in revealing that from the second claim on the stance annotations of this essay were incorrect.

2.6 Discussion

One major design challenge, we were faced with was the orientation of the graph view of an essay in the ArguLines and the AUOT. While we, as text visualization researchers, wanted to lay out the nodes representing argument units from top to bottom and the argument hierarchy from left to right following the idea of a table of contents or indented tree, our experts in computational argumentation insisted on laying out the argument units from left to right following the reading direction and the argument hierarchy from top to bottom to encode argument depth following the layout of the commonly used node-link diagrams. While it meant giving up the possibility to align the actual text with the corresponding visual elements in the graph visualization and adding another step to the construction of both the AUOT and the ArguLines (Fig. 2.3), we could not convince our experts and implemented our visualizations following their mental model.

We used a corpus consisting of 402 essays which is the largest corpus available that has a detailed markup and categorization of argument structures and contains even details about the annotation process. There are other corpora that provide nearly the same annotation detail (see Section 2.3), but they are even smaller – probably due to the annotation effort needed to create high-quality argument structure annotations. While the used corpus is somewhat a sweet spot for our visualization techniques, our system is clearly limited with respect to the size of the corpus, the total number of argument units per essay, and the maximal depth of the argument trees. For much larger corpora, the scrollable overview of the ArguLines poses a limitation while the AUOT could easily aggregate many more than 402 essays. However, if the corpus is very inhomogeneous, many nodes could be only sparsely populated. Very long essays with many argument units will lead to quite small nodes in the AOUT and are also a challenge for the ArguLine visualization. However, interactively limiting the depth of the argument units and pruning or replacing subtrees with appropriate glyphs could make the system quite scalable.

There was some discussion about the countables that were appreciated and used by two of our experts since they were very interested in outliers with an unconventional argument structure, but disliked by the third who focused more on patterns occurring in larger subsets of essays. His point was comparing the countables of one node with the bars inside another node is barely possible, as it would require scaling the bars to numbers by using the size of the selected subset of the corpus. However, the countables could be just ignored, turned off or the respective nodes pruned since those nodes would appear to be empty otherwise. In general, using different encodings in a single visualization has certainly usability issues but our experts immediately understood the concept and at two of the three made good use of it.

Unfortunately, the countables do no scale well to a larger number of essays. They are at their best within the subitizing range [70] within which the number of items is known at a glance without conscious counting; so, up to 4 pro and 4 con countables per node. Beyond that range, the number of countables is limited by the space restrictions of the containing node.

Depending on the maximum depth of the depicted argument structures, a maximum of about 20 countables can be displayed without becoming visually too close to the bar representation. While being sufficient for the AOUT visualization of a subset of the 402 essays of our corpus, it does not scale to much larger corpora. A possible solution to the problem might be the introduction of aggregated countables that represent more than one essay at once. However, having three or more different visual encodings will require some cognitive effort to derive and compare the corresponding quantities.

2.7 Conclusion and Future Work

We presented a visual analytics system for exploring, analyzing and comparing argument structures in essay corpora. The ArguLines list serves as an overview providing a short structural summary of both the individual documents and the entire corpus. Different orderings of the list ease the identification and analysis of subsets and distributions of argument units inside the corpus. These subsets can be analyzed in detail in the aggregated structure view using the Argument Unit Occurrence Tree. Cross-count ribbons revealing the evolution of stance between argument units, subnodes indicating the relative positions of claims with respect to their premises and hierarchical histograms enable detailed analysis of the differences and similarities of different subsets of documents. The document view provides examples of scholarly work and the necessary textual basis to verify conclusions from the other views as well as the annotation process. Our expert reviews revealed detailed and unknown information about the argument structures of the essays including the use of major claims and their relation to supporting and attacking claims and premises, cases of myside bias, the use of the thesis-antithesis pattern and the detection of outliers.

Using our system, we discussed the analysis of structure, expressed by hierarchical relations, the distribution of pro and con arguments and their order in a document. This already profound analysis and discussion can be further extended. One possible direction could be comparing the typical argument structures of authors with different cultural backgrounds. Does a typical Chinese argumentation look different than an American or Russian? In which way do they differ and why? Similarly, one can study argument structures under the aspect of different topics. Do certain topics entail a certain argumentation strategy? Which topics do, which do not?

In order to allow a deeper analysis of the semantics within an argumentation, the system can also be extended to not only depict stance, but also the function an argument unit has within the argumentation: examples, conclusions, restatements, pieces of evidence, counter-argument, etc. The extended classification would allow studying effective strategies and common fallacies. Encoding established schemes, e.g. arguments from examples, expert opinion or evidence, as glyphs to summarize argument structures of long, rich argumentative texts and to examine them from a different perspective is also an exciting challenge. For a quicker assessment of diverging opinions of annotators during the annotation or the confidence level of an automatic structure classifier, the ArguLines could be extended to show uncertainty information, e.g. by distorting, blurring or morphing the glyphs accordingly.

The application of the system is not restricted to the study of argument structures for (computational) linguistics. One promising direction is visual tutoring in the humanities or in journalism in order to teach the different styles of argumentation, their pros and cons. With future improvements of automatic argumentation recognition, it would be even possible to import the essays of students at an instant and comparing them with best practices in class.

Potential further use cases came up during the expert review sessions, such as visually evaluating classification results by selecting correctly and wrongly classified samples as two separate subsets in our system. This way, the classification process gets more transparent and can be corrected to yield better results. Another suggestion was compiling special purpose corpora from samples of existing corpora. The analyst would set some properties necessary for the planned research – such as: "there must be a counter argument for each argument" or "the text needs to include at least three claims that each have at least two premises" – to create a new corpus from all matching documents. We believe that these and other use cases will drive the further development of our system for visual argumentation analysis in essay corpora.

3

Comment Analyzer: A Tool for Analyzing Comment Sets and Thread Structures of News Articles

This chapter reports on joint work with Patrick Riehmann and Bernd Fröhlich at Bauhaus-Universität Weimar and Ines Engelmann and Hanna Ramezani at Friedrich-Schiller University Jena. It was published in IEEE Transactions on Visualization and Computer Graphics.
This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 16DHBKI083 (project SKILL).

Abstract

The lack of visually guided data exploration tools limits the scope of research questions communication scientists are able to study. The Comment Analyzer steps in where traditional statistical tools fail when it comes to researching the commenting behavior of news article readers. The basis of such an analysis are comment-thread corpora in which comments are tagged with various deliberative quality indicators as well as political stance. Our analysis tool provides a visual querying system for the exploration and analysis of such corpora and allows social scientists to gain insights into the distributions and relations between comment attributes, the homogeneity of thread sets, frequent thread structures and changes in comment qualities over the course of a single but in particular of multiple threads at once. We developed the tool in close collaboration with com-

munication scientists in a user-centered approach. The system has proven its utility in thorough reviews with the communication scientists, by corroborating existing findings in the literature but particularly by provoking and answering new research questions. Final reviews with five independent experts confirmed these observations and revealed the potential of the Comment Analyzer for other datasets currently being created and analyzed in the communication sciences.

3.1 Introduction

Most online news platforms such as the *Guardian*, *The New York Times* or the German *Sueddeutsche online* allow readers to discuss articles in comment sections. Readers can comment directly on the news article or respond to others' comments. The resulting threads sometimes consist of many individual comments, yet often only of a few. Under which conditions long and/or wide threads emerge, is a research question that communication scientists investigate [3]. To this end, they collect a large number of comment threads – vast forests of tree structures – and enrich the corpus with attributes for each comment, e.g., deliberative indicators (e.g., justifying or polite statements, proposals of solutions), ideological orientation and others [62, 96].

We worked with communication scientists who collected a dataset of about 7 300 threads formed by more than 14 000 comments to 160 online newspaper articles from nine German news platforms. Each thread is modeled as a tree, where nodes represent comments and links the reply relation. The comments were manually tagged with 20 binary deliberative indicators and categorical metadata (e.g., ideological orientation), which are the node attributes. Research questions to be investigated using this corpus range from "*Which political-ideological views of the audience can be identified when comparing the news platforms? Does the audience tend to reflect the political orientation of the news website or does it tend to take an oppositional position?*" to "*Can types of threads be distinguished according to the length and development of deliberative quality?*". Such questions were traditionally approached by the communication scientists with SPSS or Excel. However, statistics software is not built for structural analysis of thousands of trees.

In information visualization terminology the main challenge is to display, analyze and compare large sets (forests) of trees, for instance for contrasting the articles of two news platforms, with each article having a set of large trees of comments. Thus, we need to be able to find, investigate and com-

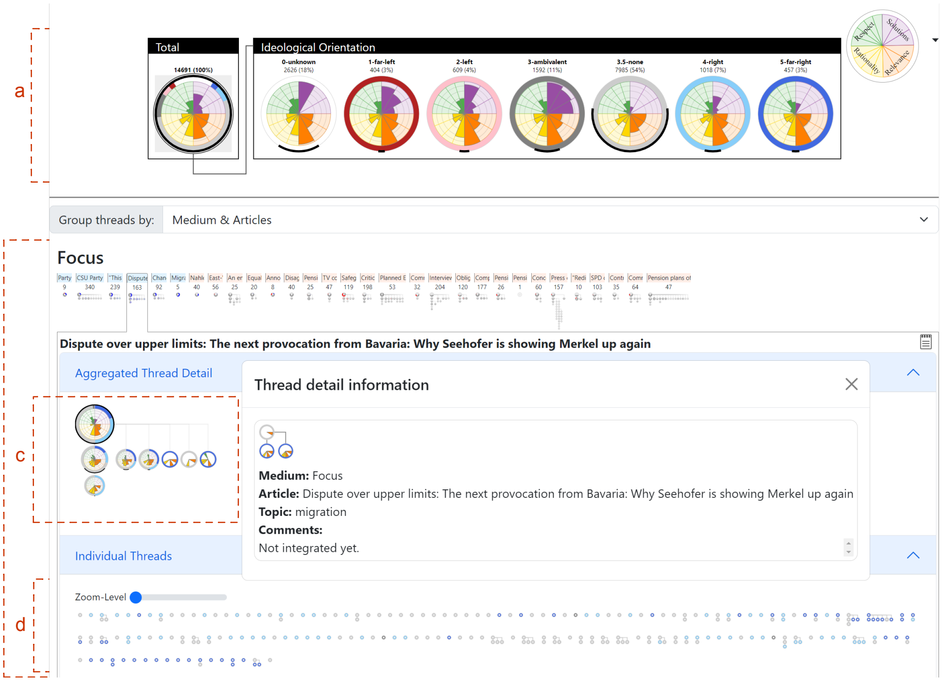


Figure 3.1: The Comment Analyzer during an analysis process: In the visual set-shaping component (a), the entire dataset has been split into subsets by ideological orientation, revealing the differences in the individual deliberative indicators. Each individual Comment Glyph shows the ideological orientation and the 20 deliberative indicators grouped by four deliberative dimensions (upper right). The set in the middle, which represents ambivalent comments, has been selected as the dataset for the structure analysis component below and therefore has a light gray background. In the structure analysis component (b), the aggregation detail view has been opened showing the Union Tree (c) and the individual threads (d). One of these threads has been selected for detailed inspection in the thread detail view.

pare node attributes, patterns of tree structures and subsets of trees. This requires the comparison of thousands of individual trees. However, displaying each tree individually can result in a cluttered and incomprehensible visualization. A solution would be to aggregate or superimpose large subsets of trees, a challenge that has rarely been addressed in research. Closely related is only Kiesel et al. [74], which is limited to binary node attributes and a single a priori aggregation criterion.

Thus, we introduce the Comment Analyzer (Figure 3.1), a visual analysis tool that enables the visual exploration of highly complex comment tree corpora. Our research provides the following contributions:

(1) **The Union Tree,** a flexible tree aggregation technique with multiple aggregation criteria, means to filter, analyze and compare the structures of aggregates as well as the relations of node attributes. It allows commu-

nication scientists to draw conclusions about common shapes of comment trees and sequences of deliberative indicators associated with the temporal development of a discussion as well as the comparison of attributes and structures of different news platforms, individual articles and article topics.

(2) **The Comment Glyph**, a carefully designed multi-attribute glyph that serves as representative of comment sets and also as a node representation for Union Trees. It visually summarizes the 20 binary deliberative indicators in a flower glyph like representation and the categorical attribute ideological orientation of the represented comments as a donut chart enclosing the flower glyph. It is tailored for semantic zooming by design and interactions with the Comment Glyph allow for filtering of the Union Trees. Moreover, the Comment Glyph also works independently of a Union Tree. Starting with the entire corpus, the social scientist is able to visually shape sets of comments by recursively splitting or merging by e.g., ideological orientation, news platform, comment length or position in reply chain. This hierarchical splitting technique keeps visually track of users' **set shaping** steps and allows the comparison of attribute distributions across all levels of the interaction sequence.

(3) **The interactive visual scrapbook** provides a persistent record of analysis results. It contributes with interactive data story elements – fully functional copies of Union Trees – that go beyond the usual functionality in notebooks. The analyst can further shape any tree set that has been stored in the notebook as a Union Tree, so earlier lines of research are not only documented but can also be refined and extended.

(4) **Reviews with two associated experts and five independent experts** demonstrated the usefulness of the Union Tree, the Comment Glyph and the interactions provided by the Comment Analyzer. During the review session, current research questions from the domain were addressed and – for the most part – solved (see Table 3.2) resulting in novel domain findings. Also, the potential of using our techniques for other datasets in communication science was discussed.

Overall, the Comment Analyzer integrates these contributions in an interactive system for domain experts. Examples of the analysis workflow can be seen in our video [73].

The rest of the paper is organized as follows: After reviewing related work, the paper will introduce the dataset and motivate the tasks the Comment Analyzer was designed to support. Following the motivation, the development process and actual design are discussed. The domain findings with the associated experts, the evaluation by independent experts and a discussion of the generalization and scalability of the approach conclude the paper.

3.2 Related Work

Tree aggregation, visualization and text. Regular tree and graph visualizations are used to convey relationships in communication (Internet Newsgroups [146], (Multi)ConVis [58, 60], Wikum [161]) or words (WordTree [93], PhraseNets [35], Netspeak Wordgraph [118, 119]). See also Yousef and Jänicke's [159] survey on text alignment, the Dagstuhl report on text visualization [21] or the Text Visualization Browser [66]. Comparing a huge number of trees is not easy. The *DAViewer* [162] compares discourse trees with interactive dendrograms, while Lahmar and Herschel [61] aggregate provenance traces. Liu et al. [89], Bremm et al. [12] and Munzner et al. [100] contrast phylogenetic trees as small multiple or through aggregation. Graham and Kennedy [45] surveyed tree visualizations, stating that aggregation – as our Union Tree – is the most space efficient. Kiesel et al. [74] is closely related, aggregating argument structures of essays, but limited to binary attributes and a fixed aggregation criterion. PansyTree [24] overlays nodes petal-like but neither the differences in the innate tree structures nor the node ordering have been studied.

Glyph design and sentiment. Similar to other designs [38, 41, 156], our work relies on round glyphs due to their aesthetics and perceived closure (see Borgo et al. [10] for general rules). Designs for multivariate data are the star glyph [40, 105] and flower glyph [16, 72, 145], both encode values by length. Part of our Comment Glyph resembles a flower glyph, but shows two hierarchical levels of grouped attributes which allows for semantic scaling. While StanceXplore [94], (Multi)ConVis [58, 60] and DoSVis [85] dedicate separate views for stance visualization, the StreamExplorer [156] uses a pie layout. Similar to Xu et al. [158], our Comment Glyph employs a donut chart to depict stance. Skau and Kosara [131] found that donut charts and pie charts work similar. Kucher et al. [85] gives an overview about stance/sentiment.

Set shaping. Our set shaping with the independent Comment Glyph allows splitting sets subsequently resembling a decision tree. While Yang et al. [152] chose Sankey diagrams, we adopt a node-link layout, like Castro and Bertini [31] and PaintingClass [137], to depict the decision hierarchy. In contrast to all mentioned works, our layout is space optimized by displaying each split of a set as one group of subsets that shares one link to connect to the original set. OnSet [125] allows direct manipulation of sets through set operations. We also allow AND and OR operations implicitly, yet not a combination.

Discourse analysis. Our Union Trees summarize the discourse of multiple comment threads. Single multi-party discourses have been studied by El-Assady et al. [28, 29] and ConVisIT [59]. Topical development and deliberative quality over time can be explored with NEREx [28] or Lexical Episode Plots [42], patterns in a discussion can be found in Jentner et al. [64] or the topic-space of a discussion explored with ConToVi [27]. The iForum [37] allows discovering temporal patterns in multiple online discourses. We extend these by combining structural and attribute-related aggregation of thread sets.

Interactive visual scrapbook. Computational notebooks contribute to the reproducibility of interim research results [83, 155]. Our scrapbook stores fully functional Unit Tree representations of intermediate results during the analysis. Scientists can interactively document, analyze, refine and compare these intermediate results during their research.

Design approach. For development, we used a bottom-up approach and showed individual tree structures to foster the design of a meaningful visual aggregation. Lee et al. [9] and van Ham and Perer [144] used a similar approach for large graph exploration. Luciani et al. [92] showed the volume details first. And even though our final system adheres to Shneiderman's top-down mantra [130], its development would not have been possible without the first bottom-up design.

Overall, our work extends the tree aggregation approach by Kiesel et al. [74] with a more sophisticated node representation and extended set shaping facilities. The flower glyph [72] was redesigned and further metadata and semantic zoom capabilities were added. The scrapbook provides versatile features for interactive visual exploration and annotation of comment tree corpora in communication science.

3.3 Corpus, Requirements and Tasks

Our experts in empirical communication science have been working for years on behavior in comment sections of online articles. They collected a corpus of 14 691 comments containing 7 263 comment threads from 160 online news articles on the nine largest news platforms in Germany about either of two topics: migration and retirement provision (pensions). With trained student assistants our experts assessed each of the comments and tagged them individually with the categorical attribute ideological orientation and 20 binary deliberative indicators (see Table 3.1). Reply relations were also captured, resulting in a forest of tree structures per news article. Furthermore, the news articles were tagged with the publication date, news platform, topic and the sequence of comment threads. To make the work more easily accessible to the readers of this paper, the titles of the news articles and a few representative discussions were translated into English.

Table 3.1: Ideological orientation and the 20 deliberative indicators grouped by their respective deliberative dimensions (color-coded). Not all of them are similarly relevant in all investigations, so usually a subset is used.

Ideological Orientation: Unknown, Far-Left, Left, Ambivalent, None, Right, Far-right	
Binary Deliberative Indicators:	
Dimension: Relevance	Topic relevance, Reply, Objective Justification, Direct Reference
Dimension: Rationality	Justification, Agreement, Disagreement, Constructive, Questions, Corrections, Summary, Meta-Talk
Dimension: Respect	Empathy, No Degrading, Politeness, Questioning Sincerity, Emotion, Humor
Dimension: Solutions	Proposal of Solution, Just Solution, Very Just Solution

With SPSS and Excel, they could only study theory-derived hypotheses regarding the attributes of the comments in relation to the construction of the news platforms, the properties of the respective articles and the deliberative indicators of response comments [62, 95, 96]. The scientists were missing a more complete picture of the corpus properties, since statistical tools are not well suited for exploratory tasks. Especially, structural relations within the comment threads and the changes of the deliberative indicators along a thread could not be analyzed. Therefore, our design focuses on the exploration of the tree structures of thread sets and individual threads as well as the discourse quality and deliberative indicators of the comments in the trees. Thus, relations within and between thread (sets) and comments can be analyzed.

Especially valuable to our experts are reply-chains resulting in a very deep conversation and broad structures, where many comments reply to the same comment, as they could reveal the circumstances and deliberative indicators that cause longer and more elaborate conversations. The variety of different and rare structures requires abstraction and aggregation techniques to preserve relevant information about the deliberative indicators and ideological orientation of each comment, typically:

- **Scanning**: Browsing thread structures and the comment attributes they contain for common or uncommon patterns in the manner of the "information flaneur" defined by Dörk et al. [25]
- **Outlier Detection**: Recognizing patterns that are unusual, unexpected or contradictory to theory both in thread structures and comment attributes
- **Grouping and Aggregation**: Summarizing discussion threads, individual threads or comments into groups by article, news platform or topic
- **Type Identification**: Identifying types of threads or comments regarding their structural or deliberative attributes
- **Set Building Shaping**: Forming sets of comments (or threads) with certain deliberative indicator combinations
- **Comparison**: Finding similarities and differences between single comments/threads or comment/thread sets

These tasks also can be (sub)tasks, consecutively and alternately applied, in their process of answering research questions and validating hypotheses. The realization of these tasks is highlighted in the following sections.

3.4 Design and Evaluation Process

Design and development followed a user-centered approach alongside the needs of two domain experts. We evaluated the fully developed system (described in Section 3.5) in two steps. First, we conducted in-depth reviews with the involved experts (see Section 3.6) to validate existing findings but more importantly investigate new ones (Table 3.2). Later on, as separate stage but with a similar protocol, we did reviews with five independent experts (also communication sciences, described in Section 3.7).

During development, we did 4 preliminary reviews with our experts. The senior expert, a professor of empirical communication studies, has more than 18 years of experience, while the "junior" expert accomplished her Ph.D. about findings from the compiled data set. The findings were published in several publications [62, 95, 96]. Both never worked with our group before and approached us for diving deeper into the structural aspect of their dataset.

The final design of the Comment Analyzer follows Shneiderman's top-down mantra [130] – Overview First, Zoom and Filter, then Details on Demand. During our user-centered development process, we use a bottom-up approach displaying all individual threads and tree structures of all articles in one scrollable view to scrutinize patterns, possible groups and outliers – supporting **Scanning** and **Outlier Detection** as well as **Type Identification** tasks for our experts and hence enabling them to pose new hypotheses. These new hypotheses then formed the design of aggregation concepts for the Comment Glyphs and Union Trees. Or to put it differently: Without being very familiar with the specifics of one's data no tailored top-down approach can be designed.

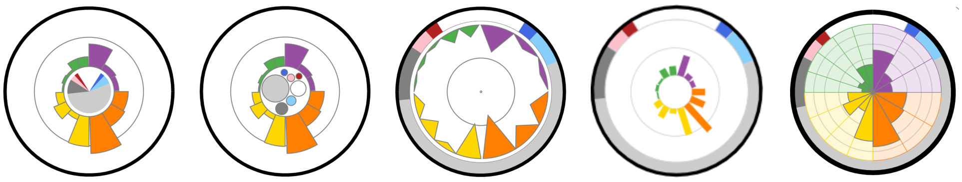
3.5 Comment Analyzer: Design and Application

The Comment Analyzer (Figure 3.1) is an interactive coordinated multi-view system composed of three main components: The visual set-shaping component, the structure analysis component and the scrapbook. Similar to the "generous interfaces" [153], we designed tangible and visually aggregable representatives for the main data items: The Comment Glyph aggregates comment sets, the Union Tree thread sets.

3.5.1 Comment Glyph

The Comment Glyph is crucial for expressing deliberation indicators of individuals comment up to the whole corpus where **Aggregation** of deliberation indicators are shown. It combines (the distributions of) ideological orientation and deliberation indicators, thus enables the analysis within and **Comparisons** between individual comments and sets.

The glyph (see iterations in Figure 3.2) conveys the presence of deliberative indicators aggregated for a user-defined selection of comments. It is organized in four dimensions of deliberative indicators (Relevance ■, Rationality ■, Respect ■, Solution ■; please see Table 3.1 for the indicators). Each dimension occupies one quadrant with individual circle slices for each deliberative indicator as a kind of radial bar. Each bar encodes the percentage of comments that have the associated deliberative indicator as the fill level of the slice. The slice positions are the same throughout the glyph instances in all components to simplify comparisons. Even though a traditional bar chart has advantages over a radial one regarding the comparison



(a) (b) (c) (d) (e)

Figure 3.2: Evolved glyph design: (**a**) Distribution of ideological orientation as pie in the center and the indicators as radial bar chart surrounding it. (**b**) Circle packing to reduce visual inferences between the pie and the bar chart. (**c**) To emphasize ideological orientation, it was moved to the outer border and combined with a spike bar chart growing from the outer ring inward to give more area to the indicators. (**d**) A radial bar chart simplifies the estimation of lengths. (**e**) Filled circle slices supported by grid rings for 25%, 50% and 75% provided the best experience.

between bars due to its more effective length encoding [154], the circular shape (principle of continuity) separates glyphs nicely from other elements. A rectangular glyph of small bar charts would require more effort to separate them from other visual elements. The fill level – or area encoding – allows a better estimation of smaller values – a desirable quality since the average frequency of most deliberative indicators is below 50% – but might lead to overestimation of larger values. The domain experts favored the area encoding over the linear one despite this disadvantage. To reduce the effect of the visual impression described in Stevens' law [134] and improve estimation accuracy, grid lines at 25%, 50% and 75% are displayed. Since an area encoding can make it difficult to read exact values despite all efforts of improvements, it is possible to switch to a linear encoding.

While the deliberative indicators are shown inside, the outer ring represents the ideological orientation(s) of comment sets. The color scheme for the ideological orientation – two levels of blue for the right stances, two levels of red for the left stances, dark gray for ambivalent, light gray for no stance and white for unknown stance – is based on the color scheme of German political parties and was also agreed with the experts.

3.5.2 Aggregating Threads as Union Trees

Due to the large variety of thread structures, we started with a bottom-up design; first showing each thread as a single left aligned tree with only a single deliberative indicator encoded as color for each node, so that the whole corpus could be initially **Scanned**. Even this basic representation impressed the experts by showing the numerous individual thread structures

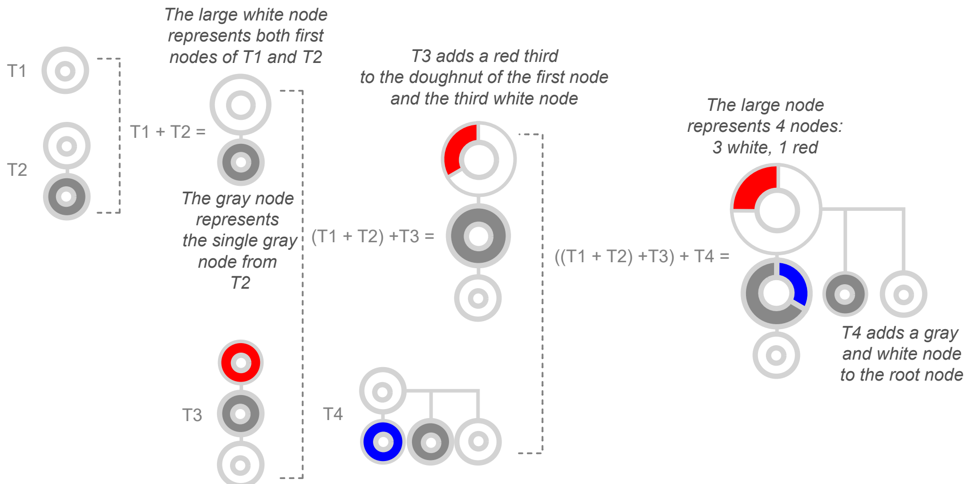


Figure 3.3: Construction of a Union Tree: The mechanism superimposes nodes while maintaining structural positions of every tree. The nodes' attributes (here shown as colored donut pieces) are aggregated. New nodes (along with their adjacent edges) were added to the Union Tree. The colors conform with the reduced color set used for small Comment Glyphs: Dark gray for ambivalent and no stance, red for left stances, blue for right stances and white for unknown stance.

paired with one selected property of their comments, which they had previously only known "by number". The sheer number of threads, however, suggested the necessity for **Grouping** and **Aggregating** the tree structures by different attributes. The Union Trees are our answer to that need.

Figure 3.3 shows how a Union Tree is constructed. Individual threads (T1-4 in Figure 3.3) – shown in a simple left aligned tree representation with simplified Comment Glyphs as nodes – visually merge into a unified structure that summarizes the nodes' attributes (aggregate at the right side of Figure 3.3). It represents the union of all individual trees by merging nodes according to their sequential position within the hierarchy, e.g., all starting comments, all first first-order replies, all second first-order replies, etc. The node frequency is encoded relative to the total number of threads in the Union Tree as size of the Glyph – the smaller, the rarer the position in the thread set. We use a logarithmic scaling of a node's area depending on its frequency relative to the root node, which improves the visibility of differences between rare nodes while keeping frequent nodes at a manageable size. In its detail representation (see upper section of Figure 3.7), the length of the thick black outer border of the respective Comment Glyph additionally encodes the node frequency; showing that only about a quarter of the individual threads actually have at least one answer – represented by the node directly underneath the root.

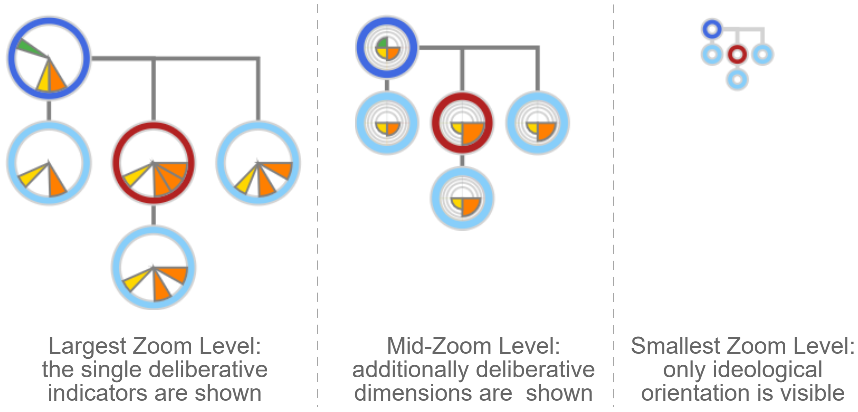


Figure 3.4: Single-Comment Glyphs at different semantic zoom levels. They are used to depict the individual threads in Figure 3.8 and T1-4 in Figure 3.3.

3.5.3 Semantically Scaling Comment Glyphs in Union Trees

The Union Trees are shown in different sizes, to which the design of the Comment Glyphs has to adapt. In the overview view (Figure 3.1b and also Figure 3.6) usually a multitude of small Union Trees need to be shown to take advantage of the given screen real estate. Thus, the Comment Glyphs show only the ideological orientation (as being the most important) either as a colored donut to represent an individual comment or as a donut chart for multiple comments (see Figure 3.4 or individual threads in Figure 3.7).

Opening a Union Tree on demand reveals its highest level of detail, using the full Comment Glyph (see Aggregated Thread Detail in Figure 3.7) for expressing multiple comments. A Glyph for an individual comment (Figure 3.4) does not show the grid for the circle segments. For a smooth appearance between these levels and to investigate dimension related research questions, we introduced a medium zoom level when zooming in and out of individual threads. In this level, the individual deliberative indicators are visually summarized by their quality dimension, facilitating down-scaling considerably without getting unreadable (see also the video [73] at 9:08).

3.5.4 Interactive Incremental Visual Set-Shaping

The set-shaping component (Figure 3.1a and 3.5) enables users to gain insight into the features of different subsets of comments. It allows **Grouping and Aggregation** and **Building Sets**, e.g., by analyzing, splitting and merging sets in various ways using three graphical elements: (1) Rectangular black bordered containers that represent **Groups** of related sets, e.g., all subsets of a split by a specific attribute; (2) Comment Glyphs depicting the subsets within one group and (3) links connecting a set with the (parent) set it was derived from.

Set Building is provided in different ways. Any set can be split into subsets regarding specific criteria: their ideological orientation, news platform, topic, reply depth or comment length. Subsequent splits result into a (sort of) decision tree with a top-aligned tree layout preserving the hierarchical nature of the splitting process with splits from one subset arranged in order underneath each other (not depicted). Semantically similar subsets of a **Group** can be merged, e.g., the far-left and left ideology in Figure 3.1. This is particularly useful for broader conclusions, for example, whether in general, left winged comments tend to contain disagreeing statements or proposals of solutions, and to further analyze the joint set by other criteria.

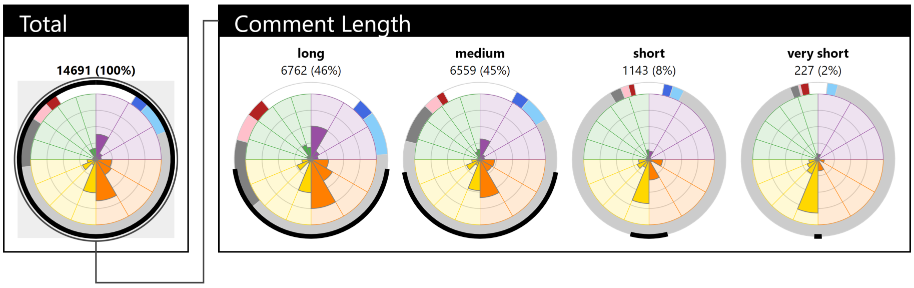
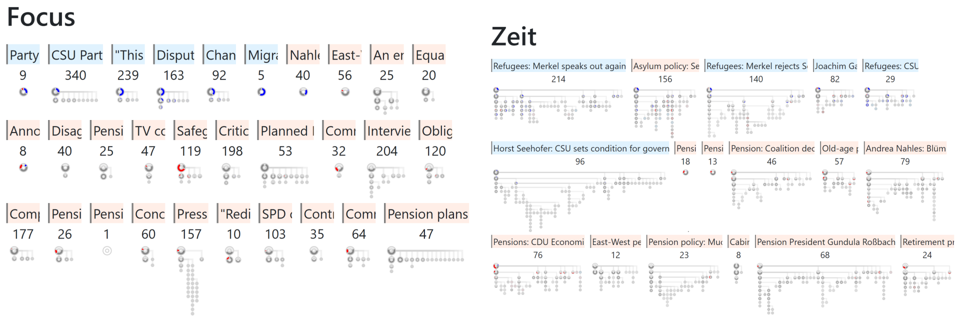


Figure 3.5: The visual set-shaping component splits the whole corpus (left) by comment length (long: > 60 words, medium: ≤ 60 words, short: ≤ 15 words, very short: ≤ 5 words). The set's sizes are written above and also encoded as length of the black outer border of each glyph. The split reveals that the more text, the more often ideological stances are detectable (the proportions of colored parts in the donut chart increase) and the more often it fits the topic (orange section next to the yellow area). Also, justifications (yellow section next to the orange area) are used more frequently in very short texts.

3.5.5 Aggregation-based Structure Analysis

Exploring and analyzing the structure of the threads and their inner evolution of a discussion characterized by the deliberative indicators and ideological orientation is one of the major features of the Comment Analyzer, which is supported by the structure analysis component (Figure 3.1b-d). It shows all threads that contain comments from a source set, defined in the set-shaping component. Relevant comments are defined by the intersection of the selected combinations of sets. Each thread in the structure analysis component must then contain at least one of these relevant comments to be displayed. These threads can then be **Grouped and Aggregated** by differ-



(a) The Union Trees of articles from the news platform Focus are usually rather flat, due to the uncomfortable comment function. The Union Trees for the migration topic (titles in blue) show a bigger right winged – blue pie slices – proportion of especially starting comments.

(b) The news platform Zeit allows replies to any comment; therefore, the Union Trees show much more elaborate structures than those seen on Focus.

Figure 3.6: The structure analysis component shows one Union Tree per article from the selected news platforms. The number above each structure denotes the number of structures summarized by the Union Tree. The text shows the start of the article's title as text and topic as background color (light-blue: migration politics, light-red: retirement provision).

ent features including news platform, articles, number of comments, depth of the thread structure and topic.

Different semantic levels help exploring the structural information of the corpus. We usually start at the smallest zoom level that shows the ideological orientation as donut charts, so multiple Union Trees can be observed and compared at once. Different **Groupings** can be selected (topics, medium, depth of discussion, etc.) to **Aggregate** the threads. Figure 3.6 shows a **Grouping** by article for two very different news platforms: Focus and Zeit. Any Union Tree can be opened to reveal the Aggregated Thread Detail view (see Figure 3.7) showing the Union Tree with detailed Comment Glyphs, as well as (on demand) all the individual threads contributing to the Union Tree unmerged in a separate view below with adjustable semantic zoom. Both views together enable up-close analysis of the distributions of the deliberative indicators and ideological orientation.

While the Union Tree clearly supports **Aggregation**, the individual threads support the **Scanning**. **Outlier Detection** and **Type Identification** is featured by both, yet on different levels. To facilitate comparison and orientation, both the Union Trees and the individual trees have a left-aligned layout that emphasizes the order in which comments were added to the thread – the further down or to the right, the newer. Interactive coordi-

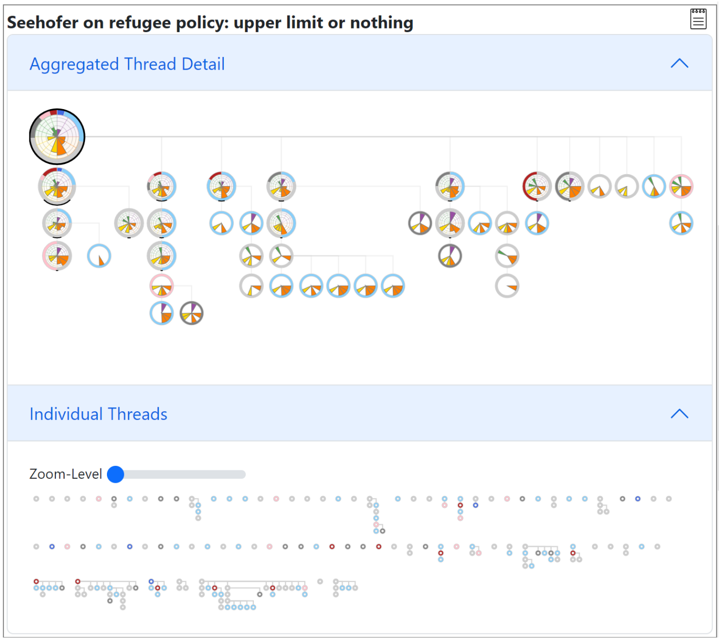


Figure 3.7: The Union Tree (Aggregated Thread Detail) in the upper part summarizes the structures and node attributes of all individual comment threads in the lower part – here all comment threads from the German news article "Seehofer about migration politics: Upper limit or nothing" – into an interactive tree. In the section below, the unmerged individual threads are shown as simple left-aligned node-link diagrams in the order of their appearance in the discussion. At the smallest zoom level, they are showing only the political orientation. See Figure 3.4 for zoom levels.

nation additionally helps when the aggregated thread detail and the individual threads are opened (Figure 3.7) by highlighting all comment nodes of the individual threads that contribute to a node in the Union Tree (and vice-versa).

For enabling further **Set Shaping** with a Union Tree, any visual component of a Comment Glyph – circle slices for each deliberative indicator, circle quadrants for quality dimensions, donut parts for the ideological orientation and the whole glyph for the node position – can filter the comments and threads to those that contain the selected feature, thereby updating both the Union Tree and the individual threads to examine the newly defined subset. Hence, a scientist can study the change in the answer probability, ideological orientation or distribution of deliberative indicators depending on the starting comment's attributes. They could study the structural and attribute differences between threads starting with a far-right or far-left (Figure 3.9) comment as discussed in the expert review section. Details about

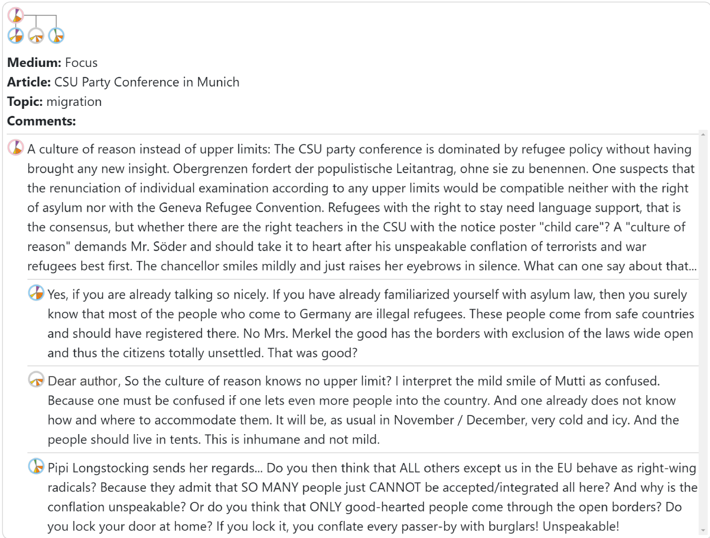


Figure 3.8: The thread detail shows full detailed Comment Glyphs, meta information and the actual texts as indented tree. The example shows a discussion about a convention of the German party CSU where upper limits of immigration were discussed. It starts with a justified and topic relevant (2 orange slices) proposition of a solution (purple) from the left-wing (light red ring) arguing against the proposed upper limits for refugee migration. The first two answers disagree (yellow slice) raising points about the applicability of the asylum law and the conditions in refugee camps. The last one contains a humorous remark relating the position of the starting comment to a character from a novel for children. The text has been translated, anonymized and truncated without removing the main point of each contribution to keep the example short.

an individual thread can be opened on demand in a separate floating view (Figure 3.8) providing the thread with all comment nodes, meta information (publisher, article, topic) and more importantly with the comment's actual text in order to validate text-related hypothesis or checking whether the previously tagged indicators are correct.

3.5.6 Interactive Scrapbook

The scrapbook component (Figure 3.9) is an interactive, notebook-like environment designed to enhance the overall workflow. It assists scientists in recording, analyzing and comparing intermediate results throughout their investigation using the Comment Analyzer. The user can store interesting tree sets – that are represented by fully functional Union Trees – and comment on the findings made within that set. Since the Union Tree is fully functional, the stored set can be drilled down further or contrasted with

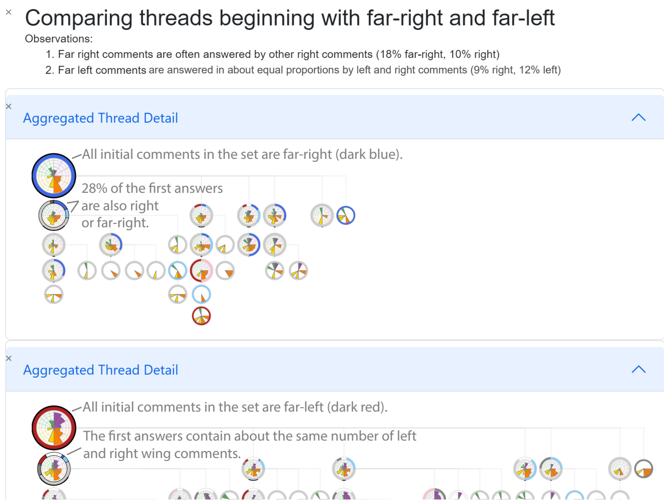


Figure 3.9: Example scrapbook during an analysis. The scientist compares the proportions of ideological orientation between threads starting far-right (dark blue) and far-left (dark red) and notes down their observations.

different filter options within the scrapbook. This way, the Union Trees in the notebook do not only document former lines of research, but can also be used as a starting point for new explorations. Storing multiple tree sets allows for detailed **Comparisons** between them. While a static view yields information about frequent structures and attributes, filtering both simultaneously, allows the researcher to look into the dynamics of attributes throughout the course of the thread. Besides **Comparing** or later reviewing tree sets, the user can take notes on observations and comparisons during the analysis. For this, they can add text sections with basic formatting (headings, lists, bold or italic texts).

The content of the scrapbook persists beyond a session, which makes it a perfect storage for intermediate results (tree sets as well as text). These recorded observations might be of great help in crafting data stories, theses or even a draft for research papers.

3.6 Domain Findings with Associated Experts

We developed the system in close collaboration with two associated experts. The development was concluded with a round of final reviews of the system and an exemplary deep analysis work flow to find answers to a list of preconceived research questions (see Table 3.2).

3.6.1 Reviews of Final Prototype

For the final reviews we followed the *pair analytics study* approach [67] with a domain expert as "pilot" and one visualization developer as "copilot" (helping out, explaining things again when necessary). The reviews consisted of (1) about 20 min of (re)introduction of the system's features, (2) a period of solving tasks and investigating the dataset and (3) an interview and feedback session. We were focusing on two things to prove the utility of our system:

(1) Can we visually corroborate findings directly that were previously described and published by our experts?

(2) More importantly, are we able to answer new research questions concerning either the deliberative indicators or structural properties alone or preferably a combination of both by employing the set-shaping and the structure analysis component?

Prior, we had our experts think of (and write down) new research questions as well as older ones to corroborate (see Table 3.2). Our experts tried to answer their questions and hypotheses during a 60-70-minute investigation phase. They were successful in all cases (see ✓in Table 3.2) except one, which would have needed a new feature in the software for aggregating predecessor and successor relationships (see also Future Work). Another question could only be answered generally, since the system lacked the means to aggregate the actual numbers of particularly selected parts of trees and subtrees. Answering each research question requires combating the basic tasks defined on page 3.3. Table 3.2 shows which ones were relevant for which question.

The results exceeded what our experts imagined. They stated having understood all visual encodings and interactions with only two minor complaints about zooming at a certain place and being forced to split based on a right click.

3.6.2 Domain-Related Findings

During the final reviews all but one research question from Table 3.2 could be answered – at least partially – resulting in novel domain findings. Please refer to Table 3.2 for more details.

P3 + P4: Validation of previous findings. In their previous studies, the experts compared single deliberative indicators and quality dimensions

Table 3.2: Research questions devised and compiled by the experts in order to proof the utility and usability of the visualization system. Except ✗, all research questions could be answered with the features of our system; two (✓) partially. The tags denote the relevant tasks for each research question.

Research Questions and Hypotheses	✓
Relevant (sub)tasks: [S]can, [B]uild sets, [A]ggregate, [C]ompare, detect [O]utliers, identify [T]ypes	
1. Previously published expert findings to be validated	
- **P1:** Comments with justification are more often followed by responses with justification than without justification. [B] [A]	✓
- **P2:** Respectful comments are more often followed by respectful responses than irreverent ones. [B] [A]	✓
- **P3:** More justifications are provided for the pension issue than for the migration issue. [B] [A] [C]	✓
- **P4:** The issue of pensions is more respectful than the issue of migration. [B] [A] [C]	✓
- **P5:** There are more response comments on pension reform than on the migration issue that (justified) disagree with the initial comments. [B] [A] [C]	✓
2. New research questions to be answered	
a) Questions about comment characteristics in comment sections	
- **N1:** Do ideologically like-minded people discuss in the respective comment sections of more left-leaning and right-/conservative-leaning news sites or ideologically right-/conservative-minded people comment on more left-leaning news sites? [B] [A] [C]	✓
- **N2:** Are there more ideological-driven comments on the migration issue than on the pension issue? – Assumption: Yes [B] [A] [C]	✓
- **N3:** Are solution-oriented comments more frequent for migration or pension reform issue? – Assumption: Pension issue [B] [A] [C]	✓
- **N4:** Does the quality of the deliberative dimensions (i.e., reason-giving/justification, civility/respect, solution orientation) show differences between the news sites examined? [B] [A] [C] [O] [T]	✓
b) Questions about thread structures in comment sections	
- **N5:** Are there fewer reply comments/reply levels on sites that don't offer visualization of the discussion thread? – Assumption: Rheinische Post, Focus and Spiegel do not offer thread visualization according to, so fewer reply comments and/or response levels are expected there. [B] [A] [C]	✓
- **N6:** For which of the two topics (migration or pension reform) do users respond more frequently with reply comments to initial user comments? — Assumption: Migration issue [B] [A] [C]	✓
c) Questions about linking comment characteristics and thread structures	
- **N7:** Do political-ideologically influenced user comments generate follow-up communication and, if so, politically disagreeing, politically agreeing or politically neutralizing? – Assumption: Disagreement is more common than agreement [B] [A]	(✓)
- **N8:** How does the quality of deliberative dimensions (i.e., rationality, respect/civility, solution orientation) change over the course of discussion threads? Does the quality of deliberative dimensions decrease, increase or tend to stay the same over the length of the thread? Contains multiple assumptions/questions, examples are:	(✓)
• **N8.1:** The quality on the deliberative dimensions remains the same in the length of the discussion. [A] [T]	✓
• **N8.2:** The quality of deliberative dimensions is better in reply comments than in initial comments, especially in Zeit Online, Welt Online and Sueddeutsche Online. [B] [A] [C]	✓
- **N9:** If the indicators change over a discussion, do they change uniformly or differently, and if differently what patterns appear? [S] [A] [O] [T]	✗

between the topics. They found the pension topic produces more respectful comments and that those comments contain the "justification" indicator more often than the migration topic. At the medium zoom level that aggregates deliberative indicators by quality dimension (see Figure 3.10a), we can verify that the pension topic shows indeed more respectful deliberative indicators (green quadrant, 34% for pension, 32% for migration). As a byproduct we found that it also shows many more proposals of solutions (purple quadrant, 56% for pension, 17% for migration, solution for **N3**) and occurrences of relevance indicators (orange quadrant, 89% for pension, 67% for migration). At the detailed level (Figure 3.10b), we can confirm that the pension topic figures slightly more justifications within its comments (yellow spike next to orange quadrant, 52% pension, 51% migration). Even better, we could derive new and more detailed insights immediately such as the striking difference in topic relevance (orange spike next to yellow quadrant), where again the pension topic provides more of (82% pension, 43% migration).

N2: Does the ideological orientation depend on topic? This detailed view also clarified a new research question about whether the distribution of ideological orientation depends on the discussed topic: The migration topic is discussed to almost a quarter by right oriented comments (light and dark blue sections of the donut), while for pensions, the right-tended comments almost do not exist.

N5: "Discussion architecture" affecting the thread structure. Another issue was how the "discussion architecture" – the features available in the discussion forum such as a visualization of the thread structures, an accessible reply button or moderation – of the news website affects the response trees of discussion threads. For this we aggregated unions trees at news platform level as shown in Figure 3.11. As expected, the platforms that, all things considered, have the worst UIs (*Huffington Post* and *Focus*) have the smallest Union Trees. Vice-versa, one would expect that the platform with the biggest Union Tree the *Zeit* would have the best user interface. Yet, this is only true in part. Generally speaking, the UI is one of the better ones, but has a fatal flaw in its reply mechanism. A reply will not be directly attached to the comment it should respond to; instead, it is placed after the latest one. The result is a rather sprawling tree with each node representing very few comments.

N7: Influence of the starting comment upon thread development. Going beyond purely structural phenomena our experts provided specific research questions only to be answered by employing both the set-shaping and the structural analysis component. Figure 3.9 shows such an exam-

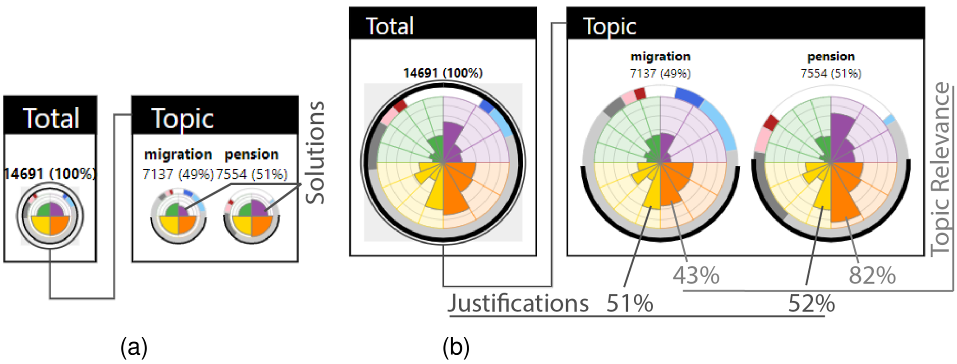


Figure 3.10: Findings about quality dimensions and deliberative indicators over two topics: (**a**) The medium zoom level shows how many comments have indicators in each of the four quality dimensions. The size difference of purple quadrants reveals more proposals of solutions to comments about pensions than about migration. (**b**) The detail level shows how many comments figure each single deliberative indicator. The glyph verifies that the pension topic contains more justifications than the migration topic. Additionally, the indicator Topic Relevance shows that comments about pensions stay significantly more on topic than comments about migration.

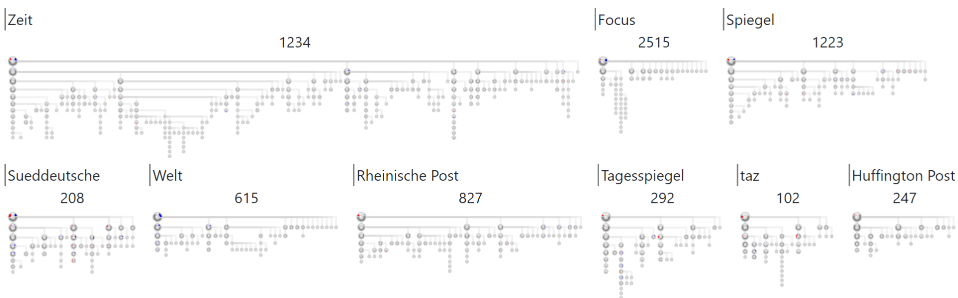


Figure 3.11: The Union Trees aggregated up to news platform clearly show the variability of the width and depth of the trees as well as the size of node areas (indicating the number of nodes existing at this position). These observations contributed in part to preliminary reasoning of our experts about the influence of well-designed discussion features provided in the UIs of the news platforms, but will also be subject to future investigation.

ple about how the starting comment influences the ideological trend in the Union Tree, particularly, if the starting comment is either far-right or far-left. Both scenarios could be shaped and filtered quickly and then put in the scrapbook for comparison. The upper part in Figure 3.9 shows that a far-right starting comment is usually immediately answered neutrally or by a "confirming" right comment (nearly 30%). Left ones only appear later and more distant. However, with a left starting comment the situation changes but we did not find it inverted (as one would expect), since the very next comment below the starting one is more balanced in terms of left and right (as well as many later ones). This observation supports our experts' assumption that right-wing commentators like to reinforce each other, while the reaction to left comments depends probably on the topic.

3.7 Reviews and Case Studies with Independent Experts

For external evaluation, we reached out to five independent experts (**IE1**-**IE5** with different institutions, ordered by date of appointment) of communication sciences, two professors (**IE2**, **IE5**), one postdoctoral fellow (**IE4**) and two senior Ph.D. candidates (**IE1**, **IE3**). None had ever seen any iteration of the Comment Analyzer before.

Again, we followed the *pair analytics study* [67] approach with a domain expert as "pilot" (operating the tool) and one visualization developer as "copilot" (helping out, explaining things again when necessary). It consisted of: (1) Introduction to the system (20 min), (2) Investigating with the system (45-60 min), (3) Interview and feedback (15 min). One of the visualization developers/designers (**VD**) performed the review with each **IE**. For the investigation part we started with simple test tasks to see whether they understood the basic concepts and were able to operate the system, such as: "*Considering ideology, which group has the most proposed solutions?*", "*How does that look like, concerning the two major topics?*", "*Which topic has the larger tree?*", "*What is the longest single thread in a Union Tree*" and "*What is the ideological distribution of the first replies and what is the distribution of the second replies?*" All **IE**s were able to solve these test tasks quickly and answer them correctly.

We continued with 4 sample research questions from our associated experts – the first research question under 1, 2a), 2b) and 2c) from Table 3.2 respectively – to ensure comparability with our associated experts' review and to provide an entry point into the investigation. All **IE**s found suitable solutions to all 4 questions. Usually, after some time, the **IE**s had their own ideas of what they would like to see or investigate with the system, which is described in the next section.

3.7.1 Case Studies and Findings

IE3 was interested in **relations between ideological orientation and news platforms**. She did different splitting cascades in the Set-Shaping component: *Total* → *Medium* {*Spiegel* → *Ideological Orientation*, *Sueddeutsche* → *Ideological Orientation*, *Focus* → *Ideological Orientation*, *Welt* → *Ideological Orientation*} for comparing the respective glyphs for *unknown, far-left, left, ambivalent, none, right, far-right*. Regardless of the political bias of the news platforms, the respective glyphs for the ideological orientations were quite similar across the news platforms, which she expected differently. However, she found that the fraction of right and far-right comments expressed in the glyphs for Spiegel and Sueddeutsche were surprisingly high given their bias towards the left, whereas Taz (as expected) had very few right comments and no far-right at all. **IE5** had similar interests and did similar splits but found an aspect that **IE3** had overlooked. The general extent of deliberative indicators was the smallest overall for the Welt, which she considered might be related to ideological orientation or to "drive-by" comments made without signing up as user of the news platform, which was still possible when the data was gathered.

IE4 wanted to investigate **audience polarization**, a topic she had gathered similar data of English news platforms for and was eager to compare it to their German counterparts in order to see whether her findings can be confirmed for German-speaking audiences. One splitting cascade she was investigating thoroughly was {*Total* → *Topic* {*Migration* → *Medium*, *Pension* → *Medium*}, *Total* → *Medium*}. It showed her that there was a clear difference in audience polarization between the two topics (as well as towards the comments overall) for each individual news platform that contained comments for both topics such as Focus, Spiegel, Rheinische Post, Sueddeutsche, Tagesspiegel, Welt and Zeit. For the migration topic the ratios of ideological orientations shifted with a significant increase of right and far-right comments for Sueddeutsche, Tagesspiegel, Welt and Zeit and even a drastic increase for Focus and Spiegel. The left and far-left comments also changed with the migration topic, foremost for Focus, where they almost disappeared, and for the Tagesspiegel, but in the opposite direction with an increase, especially of the far-left comments. Only the Rheinische Post kept their ratio balanced. The spotted trend is similar to observations **IE4**

made from compiled comments from English-speaking platforms, only that it took her minutes with the Comment Analyzer and not hours as with her previous workflow.

Concerning her audience polarization interest **IE4** also investigated **dynamics of comments and replies in light of political ideology** in the Structure Analysis component. There she grouped by topic and explored the migration Union Tree by filtering the glyphs of the starting comments by left and by right and putting the two filtered Union Trees in the scrapbook for comparison and further analysis. There was a clear difference in the left/right ratios of the comments replying to the starting one. After a left starting comment, the number of right comments of 1st-order-replies were notably higher compared to a right starting comment, where the 1st-order-replies were more balanced. One could even see a pattern that the 2nd-order-replies (answering the 1st-order-replies) were left to a large degree and had no right comments at all. On the other hand, for a right starting point such pattern simply did not exist, as one would expect.

IE1 was interested in **intensive conversations and what caused long threads** in the Structure Analysis component, starting with the working hypothesis that these were characterized by left and right comments taking turns sometimes along with "neutral" and "balancing" comments. A second working hypothesis was humor might also play a role. **IE1** filtered and visited longer threads (also put some in the scrapbook for comparison) and looked closer at their dynamics. Unfortunately, there was not a clear picture concerning this. Some resembled the first hypothesis, but more did not. Also, the humor idea did not hold. For other deliberative indicators, this first sample scan did not show a clear picture either. Further sessions are needed to find something in the indicators that might constitute such long conversations. A feature we envision in the Future Work section about fuzzy filtering node vicinities might help here. Even though **IE1** could not find a general cause for long threads, the interaction clearly showed how easily hypotheses can be tested with our system.

3.7.2 Feedback, Suggestions and Ideas

All **IE**s stated having understood the visual and interactions concepts. **IE4** appreciated the information density of the Comment Glyph, while all **IE**s found its grid rings particularly useful for assessing the values of indicators.

The most frequently mentioned complaint (**IE1** and **IE2**) was that the default size of the trees was too small, lack of labels for the indicators (**IE2**) and missing Undo capabilities (**IE3**). **IE1** would like to be able to split the glyphs in the Set-Shaping Component by the deliberative indicators and inverse filtering. **IE5** wanted to see the comments that were changed by the moderation team to see its influence on the structure of threads. All were fond of the scrapbook; being able to combine intermediate and final analysis results with their thoughts and notions as a way of recording the analytical process.

Regarding usage scenarios, **IE1** suggested it could be helpful for journalists to do analysis of their readers. **IE2** imagined that the tool could help her when teaching communication science to her students and that Master and Bachelor candidates could do analysis of similar data sets with the tool, especially, if they do not have a computer science background or are not yet proficient with SPSS or R.

Three of the **IE**s (**IE1**, **IE3** and **IE4**) were especially enthusiastic, since (previously unknown to the **VD**) they were working with similar data sets that fit the profile of our system. Two (**IE1**, **IE3**) were working with different sets of Reddit data that was tagged according to multiple attributes. **IE4** had compiled comment data from English news platforms. They were not tagged with deliberative indicators, but with a number of social values instead. Even better, her data set contained exact spots in each comment (words or word groups) being the reasons for the given tags. These locations could be easily included and highlighted in our thread detail (in Figure 3.8). Our experts may not be representative for all the communication sciences, but the reviews clearly show that there is plenty of interest for visually aggregating attributes for multitudes of comment trees.

3.8 Scalability and Generalization

The Comment Analyzer was designed for 20 000 comments. However, the set-based abstraction could handle much larger datasets, since the number of visual elements does not increase. This also applies for the structure aggregation in the Union Tree. Yet, the legibility of trees can suffer if the thread structures vary strongly since many rare nodes enlarge the resulting aggregation. The Comment Glyph depicts 20 indicators along with the ideological orientation, which might be the limit of the design in the used sizes.

The Comment Analyzer can be applied to other datasets that contain hierarchical items with multiple binary and one (or more) categorical attributes and have useful aggregation criteria. We provide insight into a second dataset example in the supplemental material: a corpus of hierarchical argument structures in student essays, where each item is a text passage with several attributes (20 binary human value indicators and the categorical attribute stance). Union Trees can also be used in other domains. In social sciences for argumentation structures; in biology for comparing phylogenetic trees; and in the digital humanities for matching ontologies.

3.9 Conclusion and Future Work

The Comment Analyzer is a visual analytics tool specifically designed for exploring and analyzing comment tree structures, developed in collaboration with social science experts. Our Union Tree aggregates multiple thread trees by visually superimposing multiple semantically related threads in node-aligned piles according to news platform, ideological orientation, etc. The Union Tree uses the Comment Glyph to represent the aggregated deliberative indicators within the node. The glyph is particularly designed for semantic zooming in order to adapt naturally to different zoom levels of tree sets or comment sets. Interactions with the glyph allow visual querying enabling detailed research about the homogeneity of the tree sets, frequent structures, changes in comment attributes in a single thread or their dynamics in multiple comment threads at once. As confirmed by associated and independent experts, it effectively combines aggregation techniques for multi-attribute and tree data to provide an overview of the data in a natural and intuitive hierarchical display. Found insights can be preserved in the scrapbook – a notebook like environment that can store fully operational copies of Union Trees and its applied filters as well as written notes on findings.

Although the presented analytical abilities exceed those of the used statistics software regarding tree data, there are still research questions suggested by our experts that require further development of the Comment Analyzer. One question was "Does the quality of the four deliberative dimensions decrease, increase or rather remain the same over the course of the thread?", that can be answered by looking through the thread representations as our experts did. However, a separate view that shows the evolution of the deliberative indicators alongside a Union Tree could facilitate the process and could also provide actual numbers. They further asked "Do subtrees at different positions in the comment tree exist that are characterized by similar deliberative indicator patterns?". Another view for analyzing reoccurring patterns that aggregates subtrees from different positions in the reply chain would be helpful in saving the communication scientist a lot of additional manual labor – even more than the Comment Analyzer already does.

4

Excursion: Rose Charts: Area or Length Encoding for Fill Level of Circle Sectors?

This chapter reports on joint work with Patrick Riehmann at Jönköping University and Bernd Fröhlich at Bauhaus-Universität Weimar. It was published in as short paper at EuroVis 2025 and was presented there. It received the Best Short Paper Honourable Mention Award.
This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 16DHBKI083 (project SKILL).

Abstract

This paper examines the accuracy of value estimation from the fill level in circle sectors of rose charts, a circular chart type where all sectors share the same angle and encode values through either radius or area. But which encoding yields more accurate estimates? We conducted a user study comparing different sector configurations and chart sizes, as well as the estimation errors of rose versus bar charts. Our findings indicate that both area and radius influence estimation accuracy. For values below 65%, area dominates as a visual cue, whereas for larger values, a transition to length perception can be observed. Based on these insights, we propose a transfer function to correct for average estimation errors and provide practical guidelines for the effective use of rose charts.

4.1 Introduction

Circular designs are a common sight in visualization contexts, used to depict time series data [39, 147], but also multi-attribute data [80]. A common one, the rose chart, consists of sectors with the same angle but varying radii. There are two encodings: either the respective attribute value is mapped linearly on the radius (as the majority of the found examples) or on the area of the slice (see Figure 4.1 (a) vs. (b)). This ambiguity means that viewers may misinterpret values in a rose chart if no clear indication of the encoding method is given. No guidelines exist on which encoding aligns better with human perception. Estimates could rely on radius by comparing the radius length with the length of the whole sector along its edges, or on area through part-to-whole comparisons. In fact, we would expect both to influence perception. Other factors, like sector sizes that change the sector's shape and the sector orientation that might introduce perspective distortion, might also bias the estimation. There is no question that the size of the chart affects the value estimation, but how small can you get without losing too much accuracy? As of yet, these questions remain unanswered.

Thus, we performed two studies investigating human perception and presentation factors for estimating values in rose charts. Presentation factors included (1) chart size, (2) number of sectors, and (3) sector orientation. A comparison with bar charts connects our results to known research. Our results provide clear indications that value estimation relies on both cues: first area, then transitioning to radius from about 65% on. Thus, we provide a transfer function that maps a data value to a radius along the range of all tested values (in 5-95%), which significantly improves estimation errors.

4.2 Related Work

The rose chart we are dealing with in this paper (Figure 4.1 (a)+(b)) is also called a circular bar chart, radial column chart, radial histograms [2], Coxcomb Chart, or Polar Area Chart. The most important historical example that uses proportional filling of circle sectors is the "Diagram of the Causes of Mortality in the Army in the East" by Florence Nightingale – part of a letter [101] sent to Queen Victoria on October 11, 1858. Similar visualization designs vary the shape of the sectors; sometimes they are disjointed and shaped as bars or columns [117]. Some of the sectors start right in the center (as with our version) or have a small inner circle [123]. Others have large center areas, empty or used for additional information [106]. A third

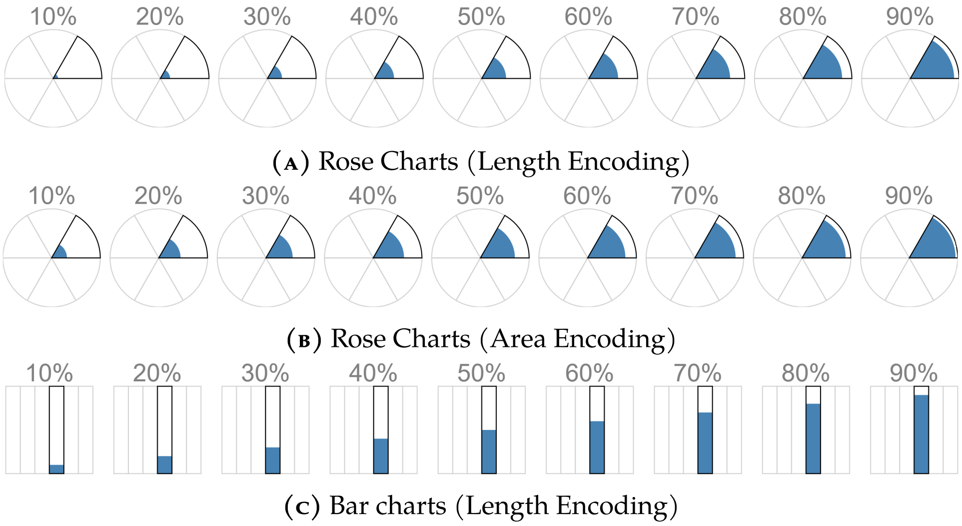


Figure 4.1: Comparison between rose charts and bar charts.

variable was the existence (or nonexistence [128]) of an outer ring expressing an outer maximum that frames the sectors.

Cleveland and McGill [20] (and Heer and Bostock [49]) showed that linear encodings outperform angle and area. This affects the reputation of pie charts, which use angle encoding and are estimated via area [81], or the arc length [131]. So far, the area is seemingly the most important cue. Pie charts are particularly suitable for part-to-whole comparisons [82].

Several studies investigate the estimation of bar charts. Diaz [26] found that boxes surrounding each bar have a positive effect. Talbot [136] investigated distances between bars, alignment, and absolute value estimation. Xiong [157] showed that viewers are best at comparing spatially close and aligned bars.

4.3 Pilot Study

The pilot study provided a first glance at the distribution of errors when estimating values in a rose chart sector, especially regarding what conditions might affect these results. The first condition is the **number of sectors** because the number of sectors defines the sectors' size and shape. We hypothesize that area recognition is active for wide sectors, while length recognition is active for narrow sectors. We chose charts with **5** and **8 sectors** that also have different symmetrical characteristics to compare charts with multiple symmetry axes against charts with only one symmetry axis. Our second condition is the **sector orientation**. Is the value estimation in

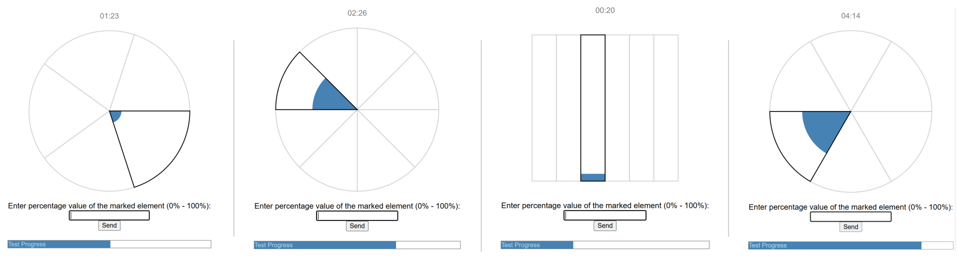


Figure 4.2: Screens as seen by the participants: two sample pages of the pilot study (left, 5 sectors and 8 sectors) and two sample pages of the main study (6 bars and 6 sector circle).

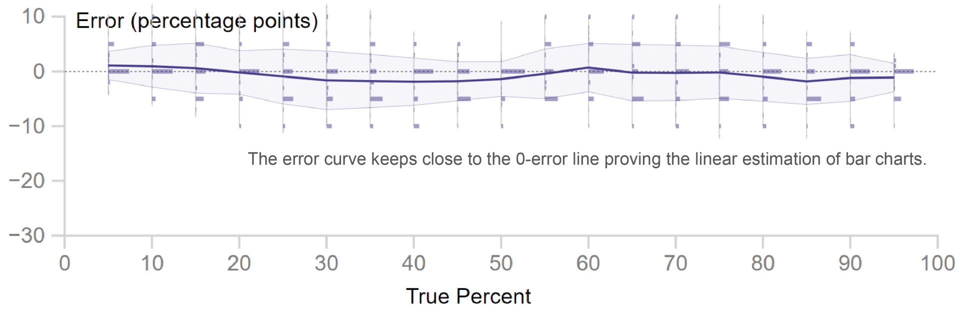
a horizontal or vertical sector different from a sector that has some other orientation? Bar chart studies like Diaz et al. [26], Cleveland and McGill [20], or Talbot et al. [136] use vertically or horizontally oriented bars, and we have not found any study that investigates bars with other rotations.

The pilot study was conducted with **7 volunteers** (colleagues, computer scientists) having at least a master's degree. It was presented as a series of websites resembling the paper card approach of Cleveland and McGill [20], but showing a single rose chart from which a single sector's value had to be estimated and entered into a text field (see Figure 4.2). The estimated values (percentages) range **from 5 to 95** in steps of 5 (19 different values) and were mapped linearly to the radii of the rose charts, e.g., we used a length encoding. The number of test cases per participant was **247 test cases** ($(5 + 8$ sectors) $\times$ 19 values). The test cases were randomized within the number of sectors. Neither the same sector nor the same value appeared twice in a row. The order of the number of sectors remained the same for all participants because no learning effect across the number of sectors was expected. The actual test pages were complemented by three pages of introduction, a survey page, and finally, one page showing the average error and time for this test run. The test pages included a timer, starting with the first task to give temporal orientation and adding a bit of pressure to prevent spending too much time on a chart and trying to "calculate" the values. A progression bar served as an additional orientation. Because the study was conducted online, actual screen densities and pixel sizes are beyond our control. Therefore, we use a calibration setup that requires adjusting a bar on the screen to the length of the shorter side of a debit/credit card, to ensure the same size of the chart on all screens.

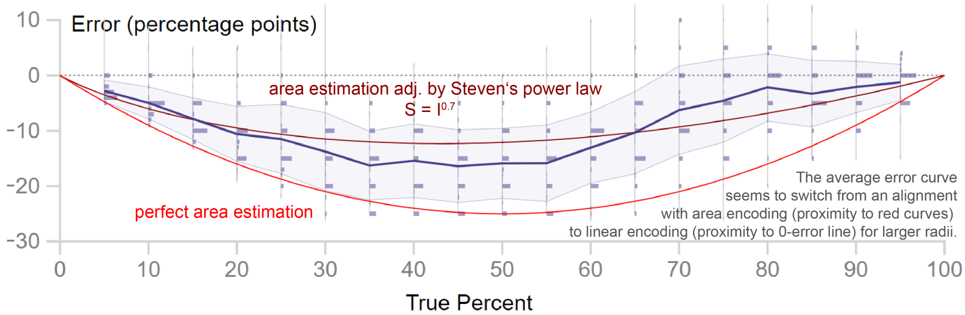
Several **findings** can be reported: **(1)** Even though the participants have not been instructed to round to **5 or 10 percent steps**, most of them did so anyway. This is in line with Talbot et al. [136]. **(2)** Using the length encoding, the average error shows a tendency to underestimate the shown values (5 sectors, M=-8.10, SD=4.10) (8 sectors, M=-7.35, SD=2.90). The absolute error grows with growing values to its largest extent between 40% and 50% and then gets smaller again for larger values. This is interesting because Talbot [136] found bar charts to get overrated for small values and underrated for large values. Rose charts do not seem to follow this rule. Using an area encoding, participants overestimated the values (5 sectors, M=9.21, SD=3.17) (8 sectors, M=10.20, SD=2.58). **(3)** A two-sided paired sample t-test between the 5-sector condition and the 8-sector condition did not reveal significant differences, t(9)=-0.99, p=0.36. Contrary to our expectations, the number of sectors does not significantly influence the estimability of rose charts – at least between 5 and 8 sectors and for our 7 test persons. **(4)** Additionally, a one-way ANOVA with repeated measures did not provide any evidence that the different orientations of the sectors affect the estimation (5-sector case, F(4,24)=0.75, p=0.57), (8-sector case, F(7,42)=0.74, p=0.64). The pilot study showed no significant influence of sector orientation or number of sectors (5-8) – this may be due to the low number of participants and should be treated with caution. However, it showed that the perceptual mechanism to estimate fill level in circle sectors works differently from that for bar charts.

4.4 Main study

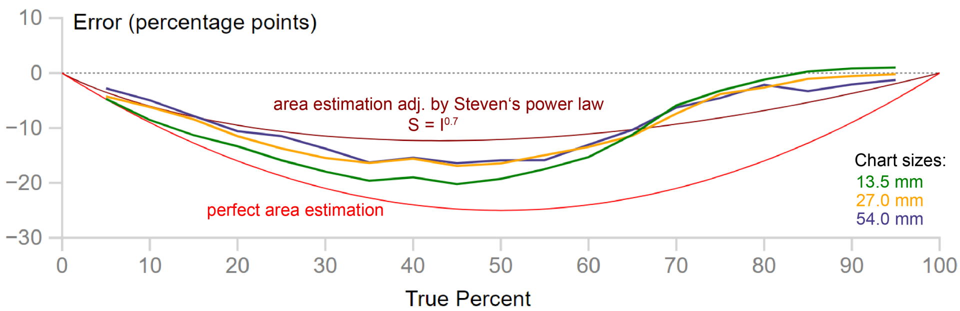
In the main study, we wanted to derive the shape of the error curve and compare the estimation errors of a rose chart to a comparable bar chart. Additionally, we are interested in the effect of the size of the chart on the estimation error. Rose charts are often used as glyphs, and as such, in smaller sizes. We wanted to determine the dependency of the estimation error on the **chart size**. Thus, we used diameters of **54.0 mm, 27.0 mm, 13.5 mm** (100%, 50%, and 25% of the shorter side of the regular debit card used for calibration). The largest size is typical for a single rose chart visualization, while the smallest size matches small glyph representations. As in the pilot study, we used values from **5 to 95 in steps of 5**. We vary the chart type between **rose and bar charts**. In our design, both diagrams consist of **6 sectors** and make optimum use of the same space. This way, the height of the bar chart is slightly less than the diameter of the rose chart, but the effective length for estimation – because the rose charts use radius and not diameter to encode the value – is almost twice as large. In a small-multiples visualization, bar charts would need the extra space surrounding the chart to



(A) Bar Chart (Size: 54.0 mm)



(B) Rose Chart (Size: 54.0 mm)



(C) Rose Charts (All sizes)

FIGURE 4.3: Histograms, average, and standard deviation of the absolute estimation errors for rose and bar charts with 6 sectors assuming a linear radius encoding. The red lines show the error of perfect area estimators (dark red: taking Stevens' power law into account, light red: without compensation).

separate themselves from each other. The closed nature of the circular rose charts does not require any additional separation.

In our within-subject study, the number of test cases per participant amounted to 114 test cases (3 sizes $\times$ 19 values $\times$ 2 chart types).

The test cases within the chart type and the chart types were randomized. We chose to vary the sector in the presentation so that participants cannot use the former values as a direct reference.

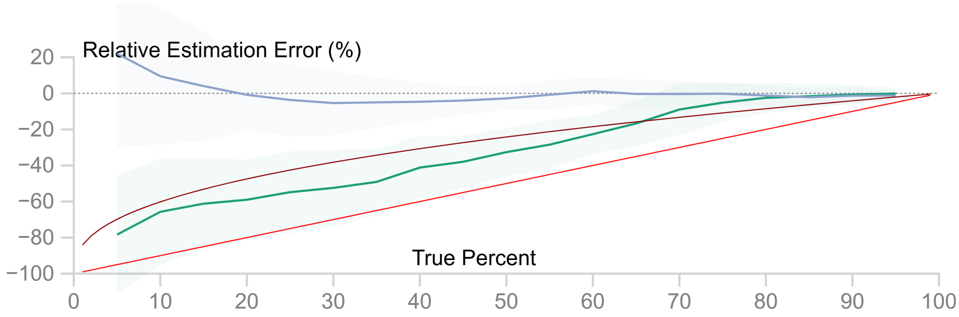


Figure 4.4: Average and standard deviation of the relative estimation errors for rose charts (green) and bar charts (blue) with 6 sectors. Again, the red lines show the error of perfect area estimators (dark red: taking Stevens' power law into account, light red: without compensation).

We reused the interface of the pilot study, except for showing two examples for both chart types (0% and 100% to not prime the participants as to which encoding is used) and adding a pause screen after half-time to avoid fatigue, as some participants reported for the pilot study. As in the pilot study, we use a calibration setup that requires adjusting a bar on the screen to the length of the shorter side of a debit/credit card, to ensure the same size of the chart on all screens. We recruited **100 participants** (mean age of 34.06 (SD = 12.20), frequency of use of visualizations (0-3) 1.06 (SD = 0.96), 64% used rose charts before) from the crowd-sourcing platform Prolific. All participants reported having normal or corrected-to-normal eyesight and had at least a high school diploma. The participants were compensated with the recommended 9 GBP per hour, amounting to 2.25 GBP for the study.

Results: Figure 4.3 (a) and (b) show the absolute estimation errors per true percent for rose and bar charts, assuming that the true value for rose charts is a linear radius-based estimate. Each plot shows the average error as a blue line, the standard deviation as a light blue area, and histograms of the measured error per true percent regarding a length encoding. A length-based estimate would show a mean error curve along the x-axis, as seen in Figure 4.3 (a) for bar charts. As in Talbot et al. [136], low values tend to be underrated by less than 3 percentage points. The error curve for rose charts shows much higher error rates, indicating that linear radius encoding does not align with human perception. We added two reference curves to the plot to determine whether the error pattern corresponds to estimations based primarily on area. The dark red curve represents the expected errors if participants perfectly estimated based on area according to Stevens' power law with an exponent of 0.7 [134]. The light red curve reflects a pure

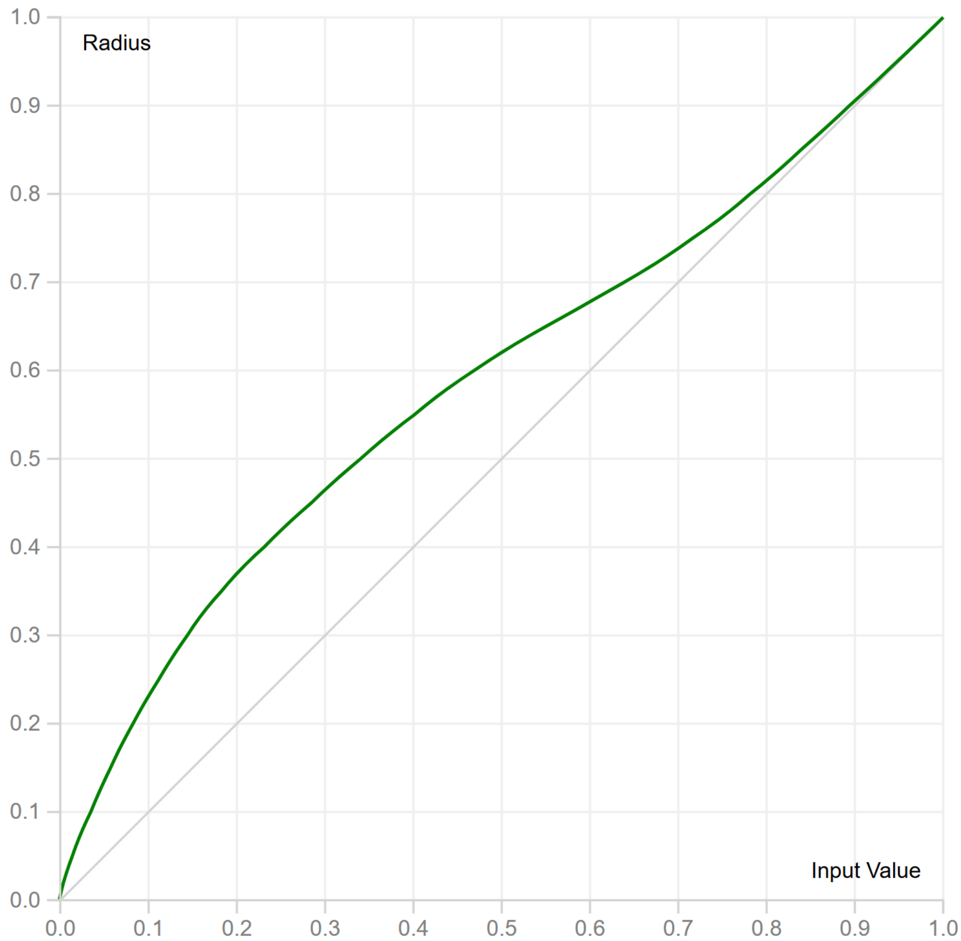


Figure 4.5: The transfer function (green line) maps input values to fragments of the maximum radius for rose charts such that the resulting estimation error is minimized.

area-based estimation without this correction. Notably, our blue **average error curve falls between these two references**. The references and the error curve show a similar shape for two-thirds of the diagram up to about 65%. The general underestimation of values from rose charts with radius coding could therefore exist because the area rather than the radius serves as a reference point for smaller values. After 65%, the slope increases, and starting with 70/75%, the error curve becomes almost linear. It seems that there is a shift in perception from area to length. Participants might have estimated the distance between the outer border of the sector and the start of the filled region for values above 75%.

The estimation error, especially between 40% and 50%, lies close to 15% for a length encoding. An area encoding with Stevens' power law applied

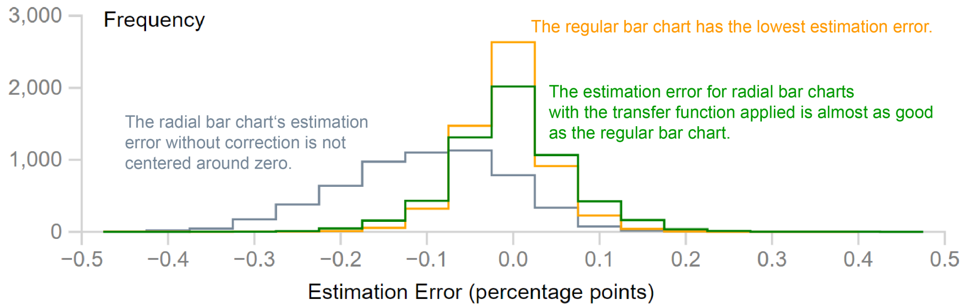


Figure 4.6: Absolute error histograms of the estimation errors before (gray) and after (green) applying the transfer function. With the transfer function applied, the estimation error becomes close but is still worse than the bar chart's estimation error (orange).

would reduce the error to about 5%. Figure 4.1 exemplifies rose and bar charts for different values and illustrates the mismatch between the encoded true percent and the perceived ones when using length encoding. If we look at the relative estimation errors made per true percent for both chart types (see Figure 4.4), the large error when using length encoding in rose charts becomes obvious: on average, participants underestimated values below 20 by 60 to 80%! An area encoding corrected by Stevens' power law reduces the error to below 20% (distance between the blue average and the dark red line). A bar chart only shows relative errors between 15% over- and 5% underestimation.

The influence of the chart size on the estimation error is evident in our data (see Figure 4.3 (c)). An ANOVA with repeated measures confirms that there are significant differences in the mean values of the sizes (rose charts, $F(2, 198)=31.4$, $p = 1.5E\text{-}12$; bar charts, $F(2, 198)=4.4$, $p=0.01$). T-tests between the individual sizes reveal that the mean error only increases significantly when reducing the size from 27.0 mm to 13.5 mm (rose charts, $t(99)=5.9$, $p=5E\text{-}08$; bar charts, $t(99)=2.9$, $p=0.005$). So it seems that a size of 27.0 mm is large enough on an average screen to fully assess the information from the plot, while a plot of 13.5 mm increases the estimation error considerably.

Error Compensation: Our comparison between rose and bar charts highlights the need for a compensation function to reduce estimation errors. While Stevens' power law for areas mitigates underestimation in radius-encoded charts (see especially Figure 4.4 (a)), it fails to capture the S-shaped curve. Instead, we model a piecewise linear filter function using 21 linear segments fitted via a sliding window approach (window size: 20%). Each point considers three to five linear values, weighted by their distance to the window center, ensuring a smooth curve. This method achieves an r^2 score of 0.96 (see Figure 4.5, green curve). The function's values are available as a CSV in the supplemental material. Figure 4.6 shows that the transformation reduces the average error to zero.

4.5 Discussion, Guidelines and Future Work

The two studies yielded several insights: (1) the number of sectors – at least between 5 and 8, and for our small sample size – does not have a significant influence on the estimation error, (2) the sector orientation – at least for 5 and 8 sectors and for our small sample size – does not have a significant influence on the estimation error, (3) the estimation error increases for chart sizes less than 27.0 mm, (4) the estimation of values from rose charts seems to be based on a mixture of area and length estimation, (5) the overall estimation error inside circle sectors using length encoding is surprisingly large. An average participant in the main study would underestimate an encoded 50% by about 15%, concluding that the chart must show only 35%! (6) An area encoding applying Stevens' power law already reduces the error significantly. (7) Our piece-wise transfer function reduces the average estimation error to zero over the entire range of values.

Regarding the size of the chart, we found that decreasing it below 27.0 mm hurt the estimation error. A follow-up study should determine the exact size at which the effect begins and how the error increases when further reducing the size. Also, we only studied the effect of 5, 6, and 8 sectors. Even though these did not show a significant difference, we would hypothesize that large numbers of sectors (>20) and very few (<4) would affect the estimations, because the visual primitives degenerate to either very thin bars or large, beveled pie slices. In our test, all sectors remained empty except for the chart's test sector. We intentionally did this to evaluate perception without distractors in this study. It is an obvious next step to add distractor sectors as conditions and evaluate their influence on the estimates.

Our results provide guidelines for the use of rose charts. The most important message is to use bar charts instead, whenever possible. Their linear length encoding makes for the best estimates. If you, however, need to use them – to match a certain aesthetic, to make the chart more memorable [11], to show patterns in a spatial embedding, or to depict a progression of time – be aware of the large estimation errors especially for small values when using a radius encoding and apply measures such as the following: (1) Implement a grid: grid lines have been shown to enhance the estimability of any plot type and will lead to considerable improvements here [6]. (2) Leave space in the center of the plot: the center of the plot produces the largest relative error; it is likely – however, not tested yet – that starting from a certain radius will reduce the estimation error for smaller values. For chart sizes above 27.0 mm, according to our study, area encoding adjusted by Stevens' power law for areas (exp=0.7) performs quite well. Our derived transfer function, shown in Figure 4.5, performs best for the investigated chart sizes, but could be fitted to specific radii to perform even better.

5
Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots

This chapter reports on joint work with Patrick Riehmann and Bernd Fröhlich at Bauhaus-Universität Weimar. It was published in Computer Graphics Forum.
This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 16DHBKI083 (project SKILL).

Abstract

This paper presents new techniques for seamlessly transitioning between parallel coordinate plots, star plots, and scatter plots. The star plot serves as a mediator visualization between parallel coordinate plots and scatter plots since it uses lines to represent data items as parallel coordinates do and can arrange axes orthogonally as used for scatter plots. The design of the transitions also motivated a new variant of the star plot, the polycurve star plot, that uses curved lines instead of straight ones and has advantages both in terms of space utilization and the detection of clusters. Furthermore, we developed a geometrically motivated method to embed scatter points from a scatter plot into star plots and parallel coordinate plots to track the transition of structural information such as clusters and correlations between the different plot types. The integration of our techniques into an interactive analysis tool for exploring multivariate data demonstrates the advantages

and utility of our approach over a multi-view approach for scatter plots and parallel coordinate plots, which we confirmed in a user study and concrete usage scenarios.

5.1 Introduction

Parallel coordinates and scatter plots are the most popular visualizations for exploring multivariate data. While parallel coordinates shine in showing relations across multiple attributes [51], scatter plots are most valuable in assessing the kind and strength of relationships and finding structural clusters [109]. The usual way to benefit from both visualizations would be to implement them in a coordinated multi-view framework. However, matching the representations of one data item in separate yet coordinated views requires the selection of its representation in one visualization and finding the corresponding highlighted representation in the other visualization, i.e., brushing and linking.

In order to enhance the cluster and correlation analysis with parallel coordinates for the finegrained analysis of argument units and explain the relationship between data patterns in both visualization types, we developed novel techniques for seamlessly transitioning between parallel coordinates and scatter plots and a new way of embedding additional cluster and correlation information into parallel coordinates. The transition is based on the observation that parallel lines intersect at infinity. Thus, parallel coordinates can be considered as a specific instance of a star plot. By tilting a pair of parallel axes against each other, the intersection point moves closer and forms the center of a star plot. The lines representing the data items are bent in an elliptic manner resulting in a star plot with curved connectors – the polycurve star plot. The polycurve star plot uses rational cubic Bézier curves instead of the usual lines to improve line tracking and provide a smooth transition to the scatter plot by shifting and the inner Bézier control points towards the scatter point and increasing their weight at the same time (Figure 5.1). As a result, the Bézier curve converges towards two lines that intersect in the scatter point. We propose to keep the lines and the scatter points during all stages of the transition. This allows us to introduce a new way to project the scatter points during the reverse transition from a scatter plot into star plots and parallel coordinates, thus adding further cluster and correlation information to the respective plots. We implemented a prototypical visualization tool for demonstrating the concept. The tool sets parallel coordinates on a rectangular layout of rails and im-

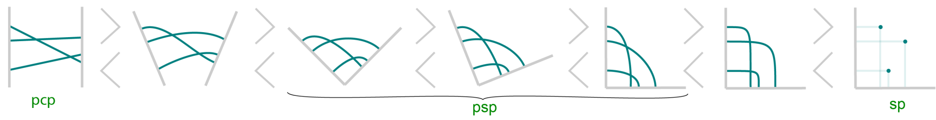


Figure 5.1: Schematic representation of the transformation from parallel coordinates (pcp) to scatter plots (sp) via polycurve star plots (psp): While the axes of the parallel coordinates plot rotate, the straight lines representing the data items bend to form the curved segments of the polycurve star plots (pcp - psp). Then rational cubic Bézier curves smoothly transition further into the scatter plot (psp - sp).

plements the described transitions on its turning points at the left and right end: first to polycurve star plots and on demand to scatter plots and vice versa. We developed the technique on well known datasets like the wine dataset [34] since their properties are known and, therefore, well suited for testing (Figure 5.2).

We were motivated by the limitations of previous attempts to combine scatter plots and parallel coordinates: Overlaying the scatter plot creates occlusion [48], integrating the intersection points of approximated regression lines produces visual clutter and difficult to interpret point patterns [164], redirecting the lines of the parallel coordinates to go through the points of an associated scatter plot distorts the information from the original parallel coordinates [160], and projecting the points onto the lines without providing a transition that can explain the emerging patterns would result in additional mental effort for the user[114]. Our work overcomes most of these limitations by the following contributions:

- A mathematically grounded technique to smoothly transition between parallel coordinates, polycurve star plots and scatter plots.
- The discussion of the properties of a polycurve star plot. It represents data items as polycurves approximating an elliptic form for orthogonal axis pairs and improves space utilization and the detection of clusters compared to regular star plots.
- The embedding of scatter points into star plots and parallel coordinate plots that are consistent with our transitions and support the visual tracking of structural information such as clusters and correlations during the transition between plot types and the analysis with parallel coordinates and star plots.
- A prototype implementation that demonstrates the seamless integration of parallel coordinates, star plots, and scatter plots into a sin-

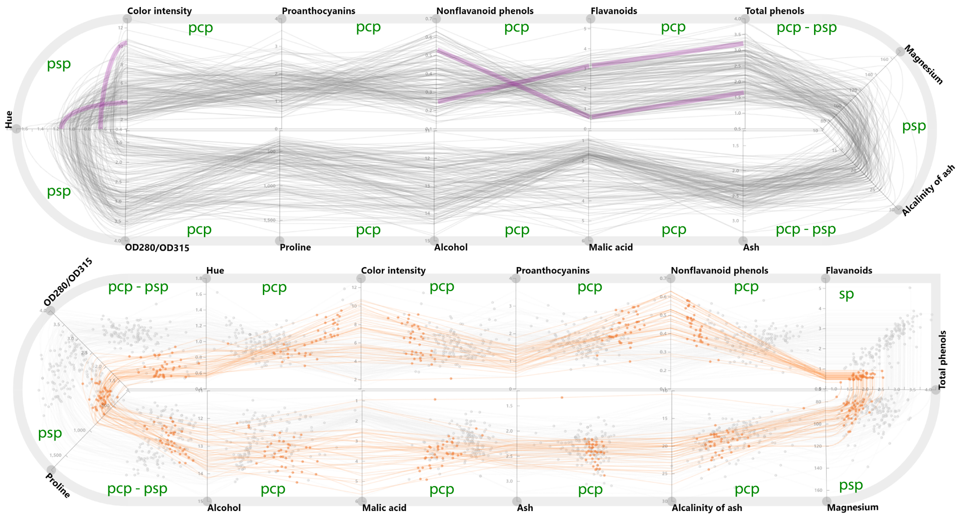


Figure 5.2: The prototypical implementation of our transition concept between parallel coordinates and scatter plots from Figure 5.1: The parallel coordinates (pcp) are set on and can be dragged along rails. At the corners the parallel coordinates smoothly become polycurve star plots (psp). When the axes are perpendicular, the user can grab and drag the corner out to transform the star plot into a scatter plot (sp). The scatter points on the parallel coordinates in the lower image give hints regarding clusters and (cor)relations. The example shows a wine dataset [34]; the purple lines mark positive and negative correlations. In the lower image, a selection was made using point selection in the scatter plot between the attributes flavanoids and total phenols which highlights the smaller cluster of less steep positive correlation.

gle view and smooth, user-controlled, geometrically motivated transitions between all three plot types.

We demonstrate the advantages and benefits of our techniques with case studies using the developed prototype and commonly known datasets and confirm their usability in a user study and the finegrained analysis of the relations between argument units and textual contents of the AAE2 corpus [133], we encountered in the last chapter.

5.2 Related Work

Scatter plots are one of the most straightforward visualization techniques but are still subject to recent research [32, 109, 129, 151]. Harrison et al. [48] showed scatter plots to be the most efficient visualization to show correlations. However, scatter plots and scatter plot matrices are less effective when comparing data points across more than two attributes at once. Additionally, they do not scale well with an increasing number of attributes. Star coordinate plots generalize scatter plots to depict more than two dimensions [87, 98] and attempt to give a holistic view of the data set. The axes are individually scaled and rotated to get the best possible view of the clusters. However, the ability to examine relationships between individual attributes is lost.

According to Harrison et al. [48], parallel coordinates [63] are the second most effective visualization for negative correlations. Positive correlations, clusters, and non-linear relations are less efficiently detectable. However, research is being conducted to improve these capabilities [163]. Heinrich and Weiskopf [51] give a comprehensive overview of the field. Holten and Wijk [57] tested several variants of parallel coordinates regarding their performance in cluster evaluation tasks and showed that only pairing the parallel coordinate plot with a scatter plot has a significant effect on cluster identification. To improve the effect, we propose not only showing scatter plots and parallel coordinates side-by-side but also smoothly integrating the scatter points directly into the parallel coordinate system.

Several attempts have been made for integrating of scatter plot information into the parallel coordinates: Holten and Wijk [57] depict a small 45 deg rotated scatter plot between two axes of the parallel coordinate plot. Yuan et al. [160] directly draw a scatter plot of a third attribute between two axes. Thereby, each line is bent to lead through the corresponding dots to link points and lines with each other. In contrast, our point projection places points onto the lines without changing the parallel coordinate itself. Likewise, the point projection of Raidou et al. [114] is applied. Their points, however, are placed based on the slope of the line in parallel coordinates. The positions of our scatter points cover more of the given space and base on the ratio of the respective coordinate values. Related but working towards a different goal are Zhou and Weiskopf [164], who integrated indexed points into the parallel coordinates system to depict local correlation. Other hybrid visualizations have been proposed. Gruendl et al. [46] seamlessly integrated time series into parallel coordinates. NodeTrix [56] combines node-link diagrams with adjacency matrices. Claessen and van Wijk [19] systematize axis-based hybrid visualizations with the ARGOI model that shows how axes are connected.

Star plots and their various variants [124] are often used in the medical field to compare patient data [14, 97]. Sangli et al. [126] extended the visualization to deal with very high dimensional data. Xie and Karki [68] extend parallel coordinates with attached star plots showing context axes to better exploit the available space in a Focus & Context view. Slightly different, Fanea et al. [30] combine star glyphs and parallel coordinates to alleviate visual clutter due to many data items. In this paper, the star plot is a mediator between parallel coordinate and scatter plot.

Animation in data visualization can be beneficial during data analysis, as discussed by Fisher [33]. Heer and Robertson [50], for example, showed a significant increase in data point tracking and change estimation using simple staged animations and a strong user preference for animation over direct transitions. Systematic analysis of animations and transitions were performed by Tominski et al. [17], who coined the term *flexible visual analytics*, by Thompson et al. [138], who explore the design space of animated graphics, and by the Gemini system [78] that proposes a grammar of animation. Ruchikachorn and Mueller [113] propose animated transitions for teaching visualization techniques. Our proposed method will follow the notion of *fluid interactions* coined by Elmquist et al. [102] and employ smooth user-controlled and naturally emerging transitions to explain the relations between data representations of parallel coordinates, star plots, and scatter plots. Bezerianos et al.[8] employ these principles in transforming different views on node link diagrams into one another. Like our work, transmogrification [13] also describes a smooth transformation of different visual representations. Their method is, however, limited to transformation through spatial distortion, while the transformation from parallel coordinates to scatter plots requires the transition from a straight line to a single point as representative of each data item.

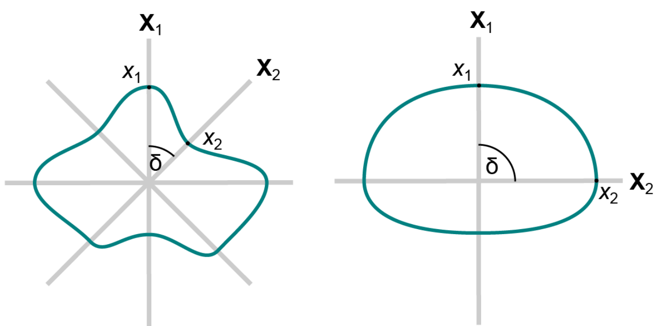
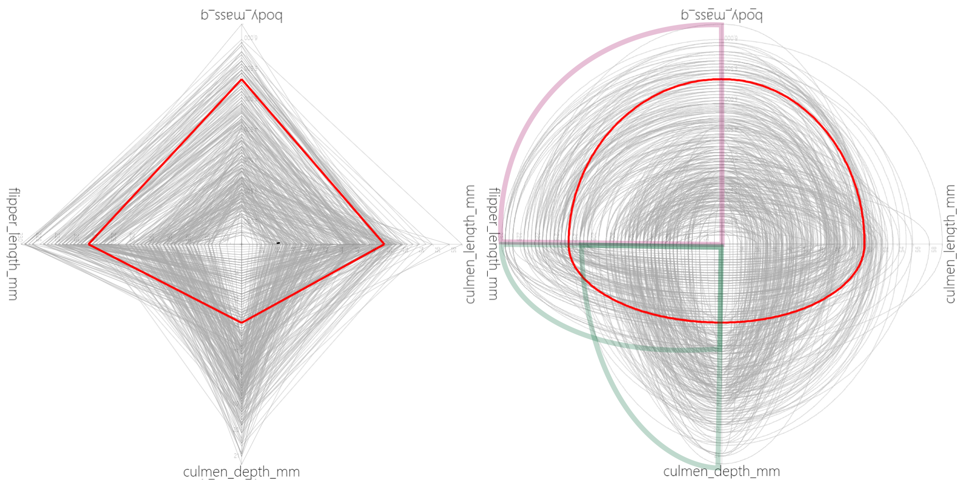


Figure 5.3: Schematic structure of the polycurve star plot: the axes $\mathbf{X}_i$ are placed at equal angles δ (left: $\delta = 45\,\mathrm{deg}$, right: $\delta = 90\,\mathrm{deg}$) around the center. The coordinates x_i define where the polycurve intersects each axis.

5.3 The need for a transition mediator: the Polycurve Star Plot

The main aspect of this paper is to define a mathematically sound, intuitively comprehensible and aesthetically appealing transition from parallel coordinates to scatter plots that works in both directions and shows the data at any stage during the transition in an interpretable way. For that the continuity of the polylines connecting the data values within the parallel coordinates has to be guaranteed while seamlessly transforming the lines towards the points of a scatter plot as shown in Figure 5.1 and described in Section 5.4. To further facilitate the overall impression of looking at a single visualization in different stages and not at two distinct visualizations shapeshifting into one another we provide a novel method for projecting the points of the scatter plot onto their respective lines in the parallel coordinates display (see Figure 5.2 and Section 5.5). This way important phenomena from the scatter plot are directly embedded into the parallel coordinate plot to begin with and all the way during its metamorphosis towards the scatter plot.

In order to ensure that smooth transition of the line representation to the scatter points, we need the starplot visualization to meet specific requirements to properly function as a mediator. For orthogonal axes pairs, we require the connecting curve to form a circle segment for pairs of equal data values; ellipse segments for unequal ones. For other angles δ between the axes, the curve segments shall retain the ellipse's property of intersecting with the axes perpendicularly, such that data items are visualized by a continuous smooth shape strengthening the connection between polycurve segments and allowing for easier tracing of polycurves (see Figure 5.3 left).



(a) standard star plot (b) polycurve star plot

Figure 5.4: Comparison of standard and polycurve star plots using the penguin data set (345 samples)[1]. The curved connectors in (b) allow for some interesting observations: In the upper left quadrant (purple mark), only few crossings of curves occur, suggesting a positive correlation. In the lower left quadrant, on the other hand, a significant number of crossings is visible. On closer inspection, one can observe that the crossover stems from two clusters of positive correlation (green marks), since within each band, only few crossings are visible. This conclusion could not be drawn from the standard star plot in (a) due to heavy overplotting issues.

A polycurve star plot using rational cubic Béziers can be constructed to meet these criteria. It employs one axis $\mathbf{X}_i$ per dimension arranged at equal angles δ around the center as regular star plots, yet with polycurves instead of straight polylines representing a data item (see Figure 5.3). Those are modeled as rational cubic Béziers, since they are able to perfectly represent ellipses, allowing us to always intersect axes orthogonally and offer a smooth way to transition to the scatter plot by increasing the weights of the inner control points. Figure 5.4 shows an example of the penguin data set [1]. Though the general concept of polycurved star plots is not entirely new (it is used for some infographics and charting tools, e.g. Microsoft Excel, ggplot, or d3js) it was never thought of as mediator visualization. Hence, our work will provide – to the best of our knowledge – the first scientific consideration of its benefits and usage as a transitional mediator between parallel coordinates and scatter plots.

Although the polycurve star plot started as a concept to smooth the transitions between parallel coordinates and scatter plots, the design has some additional advantages: (1) The design makes better use of the available space compared to a regular star plot, thereby reducing some of the overplotting issues. (2) The curves conform with the Gestalt Principles of Continuity and Closure to foster the traceability of individual data items and the Gestalt Principles of Proximity and Similarity on groups of curves to highlight patterns and clusters (Figure 5.4 (b)).

5.4 Smooth Line Transition between PCP and SP

The transition between parallel coordinates and scatter plots depends on two mathematically grounded considerations: (1) parallel coordinates can be seen as special instance of star plots with axes crossing at infinity, thus one form can be transitioned into the other by moving the intersection point of the axes (Figure 5.1 pcp-psp). (2) shifting and increasing the weights of a rational cubic Bézier curve's inner control points smoothly transforms an ellipse segment to a rectangle segment with its corner being the scatterpoint (Figure 5.1 psp-sp). The idea for this transformation was motivated by superellipses which can also transition from an elliptic segment to a rectangular segment; however, Bézier curves are easier to use since superellipses require the use of large exponents to converge towards a rectangular shape.

We uniformly denote all axes – the infinite lines defining the spatial position and angle of an attribute – as $\mathbf{X}_i$ and coordinates x_i (see Figure 5.5 and

Figure 5.3). Additionally, d denotes the distance between axes for parallel coordinates, δ is the angle between axes in the polycurve star plot.

5.4.1 From Parallel Coordinates to Polycurve Star Plots

The transition from parallel coordinates to polycurve star plots (Figure 5.1 (pcp - psp)) follows geometrical considerations: the axes of the parallel coordinates intersect at infinity. By rotating two adjacent axes left and right, respectively, the intersection point of the axes moves into the finite range. Thereby, the range of data values is also moved to remain in the visible area. Since the transformation distorts the space between two axes, the lines start to bend while the axes are rotated, resulting in ellipse segments for perpendicular axes. Thus, we reinterpret the star plot's radial layout to conform with polar coordinates: The coordinates are interpreted as radii instead of lengths connected via ellipse segments rather than straight lines.

5.4.2 From Polycurve Star Plots to Scatter Plots

The design of a transition between polycurve star plots and scatter plots relies on a rational Bézier curve's ability to represent ellipse segments and converge towards rectangle segments (see Figure 5.1 (psp - sp)). A rational Bézier curve in general is defined as:

$$B(t) = \frac{\sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i} P_i w_i}{\sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i} w_i} \tag{5.1}$$

were P_i are control points and w_i are weights. The higher the weight, the closer the resulting curve will pass by the respective control point. While a rational quadratic Bézier curve – so $n = 2$ – suffices to perfectly represent

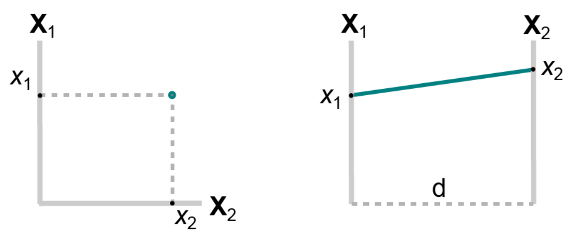
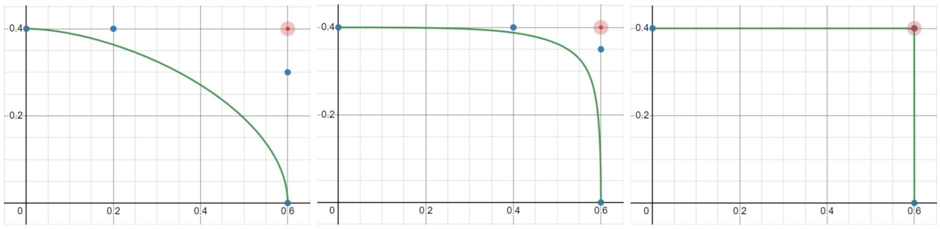


Figure 5.5: Schematic structure of scatter plot and parallel coordinates denoting the convention of variables used in the paper: coordinates x_i are situated on the axes $\mathbf{X}_i$ to construct the point (scatter plot), line (parallel coordinate) or curve (polycurve star plot). d describes the distance between two axes in parallel coordinates.



(a) initial state (b) control points move; the weight increases. (c) final rectangle

Figure 5.6: The interplay between linearly shifting the control points (blue) towards the actual scatter point (red) and exponentially increasing the weight of the control points transforms an ellipse into a rectangle segment.

an ellipse segment, we use rational cubic Bézier curves which obviously can also represent rational quadratic Bézier curves but additionally allow us to represent S-shaped curves as required for general polycurve starplots with tangential transitions across axes as shown in Figure 5.3. In order to achieve the desired effect of turning the ellipse segment into a rectangle, we not only need to increase the weights of the inner control points P_1 and P_2, but also shift them gradually towards the scatter point (Figure 5.6). For the whole transition to feel linear – as required later in our implementation – the control points are shifted linearly towards the scatter point while the weights are increased exponentially towards infinity to pull the curve into the rectangular corner.

5.5 Smooth Point Transition between PCP and SP

We introduce a new projection of scatter points onto parallel coordinates and star plots to continuously show the scatter point on the curves and lines throughout the whole transition which avoids blending in the actual scatter point of the scatter plot after the transition of the line representation has concluded. Additionally, some of the structural information that allow the scatter plot to be so effective in cluster and correlation analysis become accessible in the PCP. The projection is based on the definition of a regular scatter plot's point: A coordinate pair (x_i, x_{i+1}) places the point in the scatter plot such that it has a distance of x_i from $\mathbf{X}_{i+1}$ and a distance of x_{i+1} from $\mathbf{X}_i$. In Figure 5.5, we marked these distances with light gray dashed lines. The gray dashed lines form the rectangle segment that results from our transition described in section 5.4. When the range of data values is normalized to $[0, 1]$, the position of the point on this rectangle can be described by the ratio of the coordinate x_{i+1} (the length of the horizontal gray dashed line in Figure 5.5) to the total arc length of the rectangle (both gray dashed lines in Figure 5.5), so by the fraction f:

$$f = \frac{x_{i+1}}{x_i + x_{i+1}} \tag{5.2}$$

The points of the resulting embedding move smoothly during transitions and can be used for the analysis in polycurve star plots and parallel coordinates (Figure 5.7).

Since x_i and x_{i+1} need to be normalized to range $[0, 1]$, the fraction f for (0,0) results in a division by zero. Thus, the scatter point representing (0,0) is not defined and will not be displayed in the PCP. The nature of the transition explains the peculiarity: As the axes are pulled apart in the transition from star plot to parallel coordinates, (0,0) representing the point of contact of the two axes is also pulled apart and degenerates to a line.

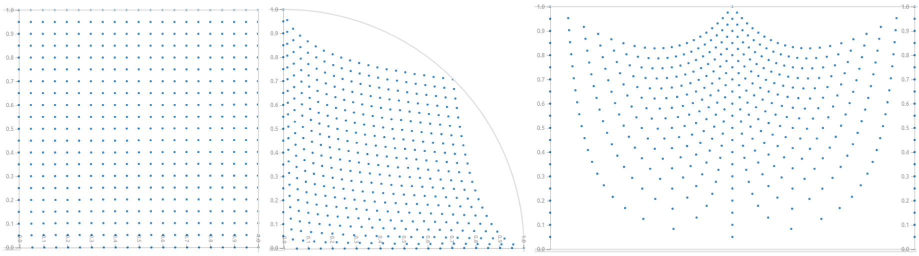
Please note that these projected points have an entirely different meaning than those proposed by Inselberg himself [63]. Inselberg's points are constructed from a line in a scatter plot and could be used for example to visualize a regression line as a summary of many single coordinates. In our point projection, each line is enhanced by one point that encodes the same information in another way, thereby enabling the user to draw refined conclusions.

5.5.1 Scatter Point Patterns in a Parallel Coordinate Plot

Geometrically, fraction f projecting a scatter point onto a parallel coordinate describes the normalized angle of the polar coordinate:

$$f = \frac{\arctan \frac{x_{i+1}}{x_i}}{\frac{\pi}{2}} \tag{5.3}$$

All points with the same ratio between x_{i+1} and x_i, so the points on a straight positively sloped line through (0,0) will have the same fraction f. Transforming a straight positively sloped line through (0,0), thus, results in a straight line parallel to the axes in parallel coordinates. Corresponding correlation patterns – global strong positive correlations – are, therefore, also vertically oriented. Since in our model of transforming the parallel coordinates to polycurve star plots (see subsection 5.4.1), we moved the intersection point from infinity to a finite position, it makes sense that all lines through (0,0) in a scatter plot are parallel to the axes in parallel coordinates. Our positioning of the scatter points is consistent with this.



(a) scatter plot (b) star plot (c) pcp

Figure 5.7: A regular grid of scatter points in (a) scatter plot, (b) polycurve star plot, and (c) parallel coordinates (PCP): each transformation warps the original grid further while still maintaining structural information, especially along the main diagonal. (0,0) is missing in the PCP, since it is spread across the lower end.

Figure 5.7 shows the distortion in an originally regular grid when transformed to parallel coordinates. Points near (0,0) are dragged apart, while points near the end of the diagonal are compressed. The embedding transforms the grid's straight columns and rows in the scatter plot to curves in the parallel coordinate system. Still, there are three straight lines in the distorted grid, all parallel to the axes: one in the center between the two axes that constitutes the diagonal and one directly on each axis. These result from grid lines originally forming straight lines through the origin that will transform into parallels to the axes as described above. The projection of the main diagonal onto a straight centered axis-parallel line is particularly useful since the projected point cloud of a positive correlation will arrange itself close to it, providing additional visual cues to an otherwise hardly detectable PCP line pattern. While the embedding does not preserve the regularity of the original grid, structural relations like neighborhoods are largely preserved so that the patterns can give valuable cues regarding the estimation of relations between attributes and the detection of clusters.

Strong positive relations translate into either parallels to the axes if the regression line intersects $(0,0)$ or into curves that tend towards parallelism on one end and intersect with one axes on the other forming smoothed L-like shapes (first row of Figure 5.8). Strong negative correlations show as U-like shapes intersecting with one or even both axes (second row of Figure 5.8).

The lower half of Figure 5.8 shows two different linear separable cluster configurations. Even though the shapes of the clusters are distorted due to the embedding, the clusters are still clearly separable since structural relationships are preserved. Although densities are distorted during the transformation, even the clusters in the bottom row can be easily told apart. So,

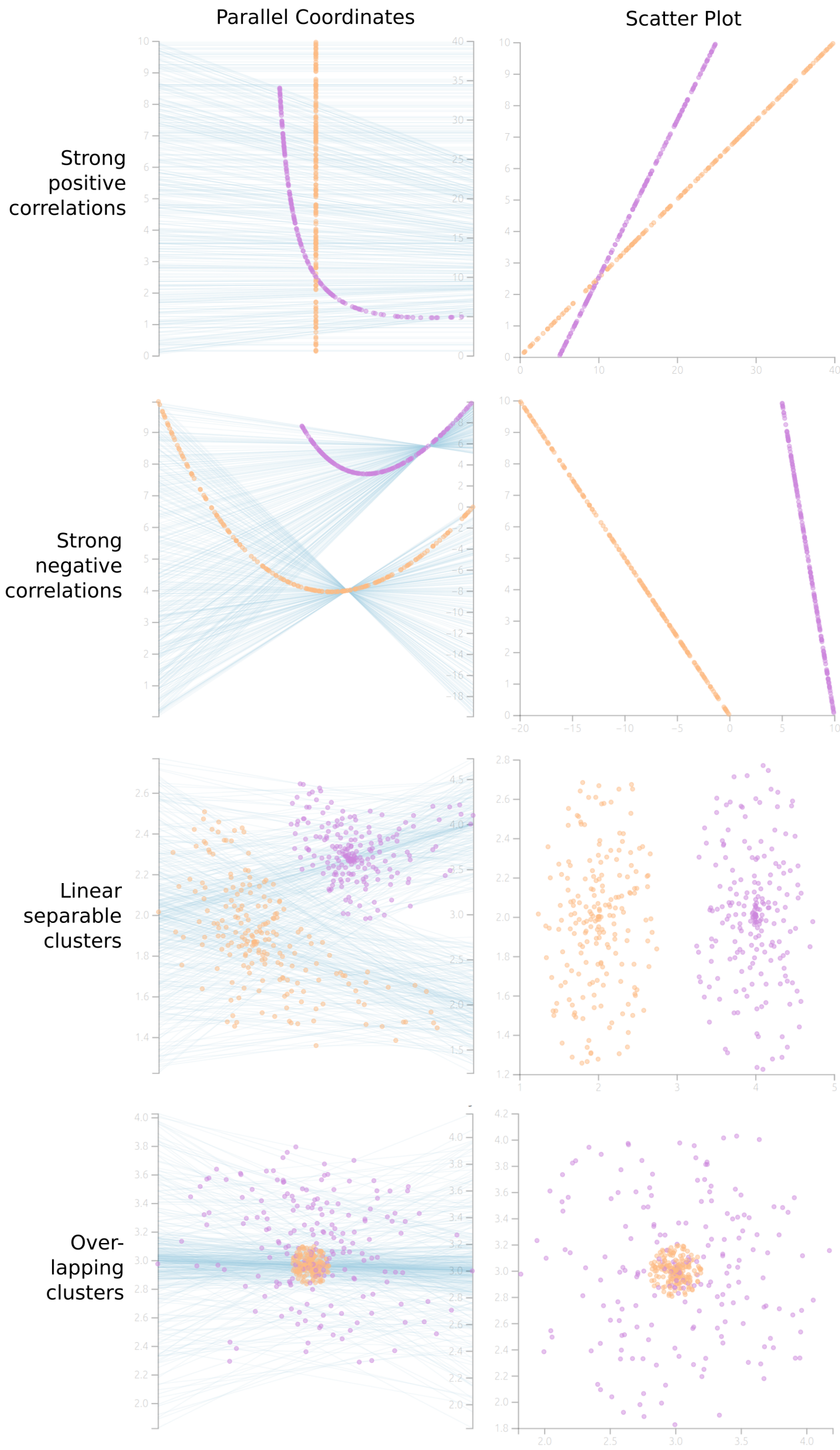


Figure 5.8: Typical correlation and cluster patterns formed by the scatter points in the parallel coordinate plot

the embedding of the scatter plot into the PCP results in an informative visual aid in analyzing cluster and correlation patters in parallel coordinates.

5.5.2 Comparison to the Point Definition of Raidou et al.

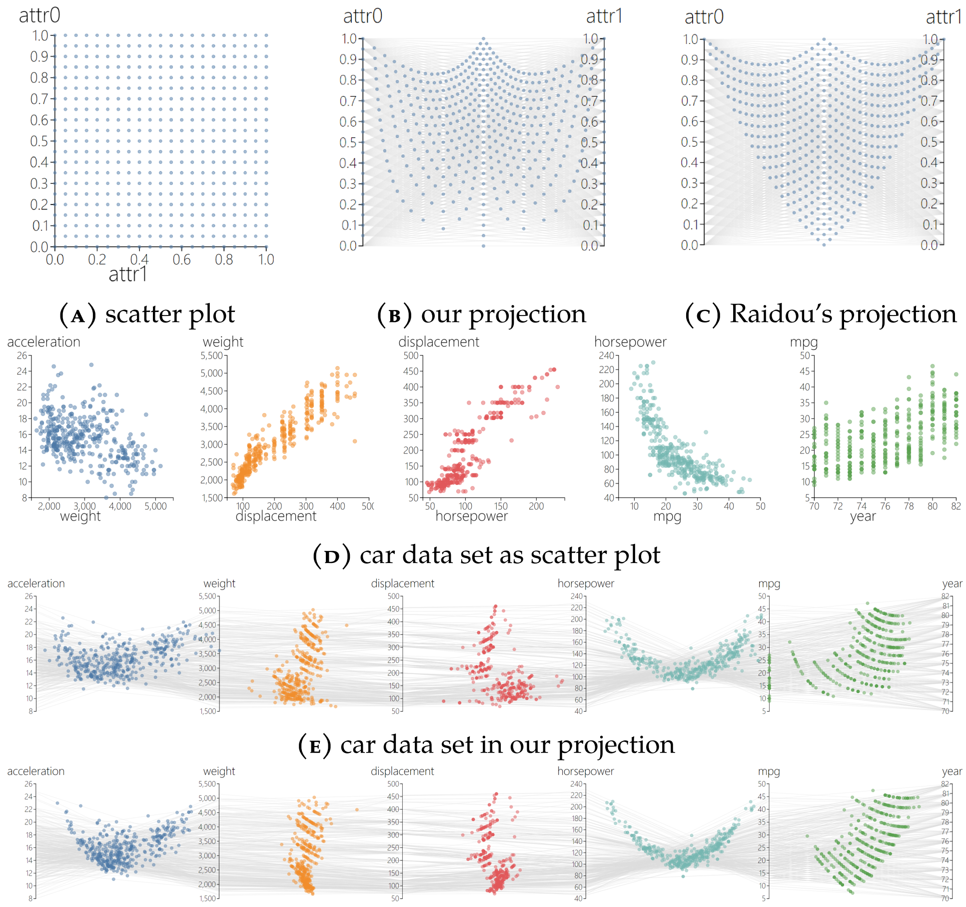
Raidou et al. [114] designed a slope-based embedding: horizontal lines get a point in the center; the points of positive-sloped lines are closer to the axis $\mathbf{X}_{i+1}$; the points of negative-sloped lines are closer to axis $\mathbf{X}_i$ ($|\mathbf{X}|$ denotes the axis length, d is assumed to be 1, Δ represents the slope):

$$\begin{aligned} x_p &= \frac{\Delta}{2|\mathbf{X}|} + \frac{1}{2} \\ y_p &= x_i + \Delta \cdot x_p \\ \Delta = x_{i+1} - x_i; -|\mathbf{X}| &\leq \Delta \leq |\mathbf{X}| \end{aligned} \tag{5.4}$$

Raidou's method generates a warped embedding of a scatter plot in the parallel coordinates plot. Their grid is more compressed but also more evenly spaced than the scheme introduced in this paper, as can be seen in Figure 5.9 (c) and (e). The tighter layout in Raidou et al. [114] introduces fewer distortions, especially in the lower half of the plot. Assuming a normalized range of values of [0,1], (0,0) is represented by one point in Raidou et al. [114]; in our scheme, (0,0) is mathematically undefined but can be interpreted as being spread across the line between the lower ends of the axes. The behaviour of points on a scatter plot's axes also differs: Our scheme projects the points on the axes as one would expect; Raidou et al. [114] place the points on a curved line between the center of the line between the lower ends of the axes and the upper end of the respective axis (green points in Figure 5.9 (b), (d), and (f)).

5.6 Interactive Analysis Tool

We designed a web-based, interactive analysis tool to demonstrate the natural interplay between parallel coordinates, polycurve star plots, and scatter plots. It implements the smooth, user-controlled transitions from parallel coordinates to scatter plots via polycurve star plots and the scatter points, the embedding of the scatter plot into the PCP. The application sets the axes of the regular parallel coordinates plot on rails, thus defining a closed path along which the whole plot can be dragged (Figure 5.2). Arcs connect each adjacent pair of straight rails. When dragging the parallel coordinate plot over an arc, the parallel coordinate representation is seamlessly transformed



(a) scatter plot (b) our projection (c) Raidou's projection

(d) car data set as scatter plot

(e) car data set in our projection

(f) car data set in Raidou's projection

Figure 5.9: Comparison of the scatter points from this paper with those from Raidou et al.[114]: Both methods transport indicators for positive (I-shaped point clouds, orange and red) and negative correlations (U-shaped point clouds, blue and cyan). The lower parts of our projections are more spread out than those of Raidou et al.[114]. Points on the axes (green) are only detectable in our scheme.

into a polycurve star plot, thereby revealing groupings of similar data items and simplifying the interpretation of the projected scatter points (see Figure 5.2). The inner arcs have configurable radii to reduce overplotting of the polycurve star plot near the center. In order to transform the polycurve star plot into a scatter plot, the user simply pulls out the outer border, which then becomes square (Figure 5.2, sp on the right turn). We decided to give the control over the transition to the scatter plot explicitly to the user instead of doing it automatically, so that the user can stop at any in-between state, better understands what happens during the transition, and to avoid introducing the interaction delay an animation introduces.

For both our testing use cases (Figure 5.2 and Figure 5.10) and the application to the analyis of argument units and textual contents (section 5.11), we chose a slim rectangular shape to optimize screen space usage. This layout creates four 90 deg angled spaces for polycurve star and scatter plots, which is the ideal configuration for both visualizations as established in subsection 5.4.2. In order to optimally exploit the available space, the parallel coordinate plot is spread around the whole rail and closed to form a band of axis pairs (or a closed path of axes according to the ARGOI model [19]). Axes can be dragged to be swapped allowing different axis configurations. Other layouts or angles are possible and might be more useful in special scenarios.

The application offers two brush types: one to select lines, the well-known standard that can be easily used to define value ranges along the axes; one to select scatter points to simplify cluster selection. Moreover, the user is able to adjust the opacities for lines and points separately. Points can be faded out when the lines should be scrutinized for correlations, or the lines can be faded into the background when analyzing the scatter plots.

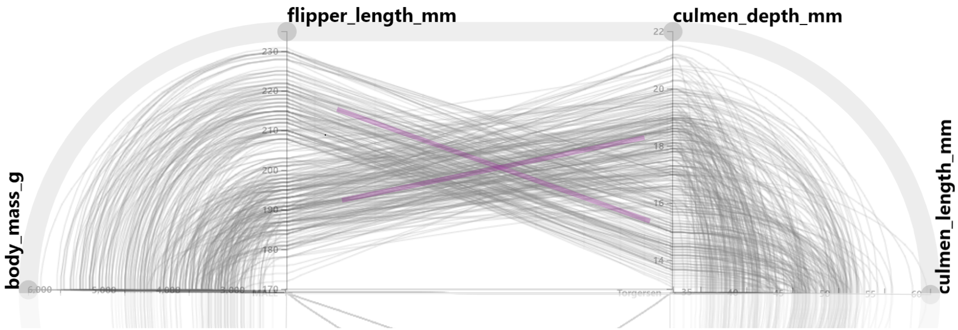
Other PCP implementations use cubic curves instead of straight lines [44, 120] for connecting the axes. Our own experiments show that cubic curves for the parallel coordinate plot make the transition appear even smoother and strengthen the traceability of the polycurves through the plot, we still decided to use the standard implementation of the parallel coordinate plot with straight lines because they convey important information for judging local or global correlation in the data which might not be as apparent if cubic curves would be used.

5.7 Proof of Concept with Known Datasets

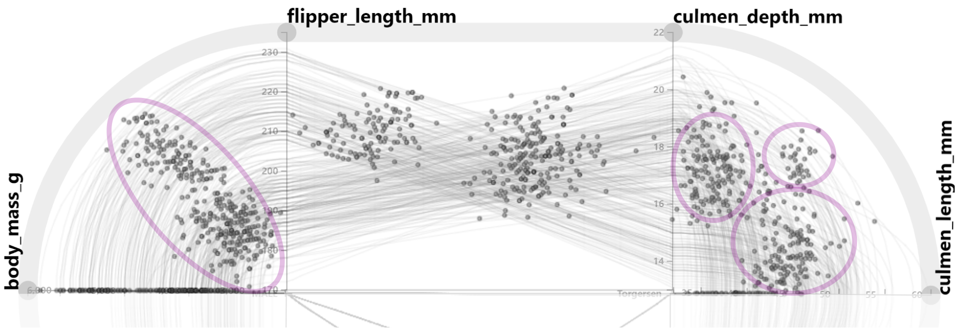
The first use case analyses a dataset of 179 wines with 13 continuous wine quality indicators (see Figure 5.2)[34]. Looking at the polycurves, there seem to be several negative correlations (e.g. *flavanoids – nonflavanoid phenols or hue – color intensity*, purple crosses in Figure 5.2) as well as a positive one between *flavanoids* and *total phenols* (Figure 5.2, purple parallels). To check whether this is indeed a positive correlation, we first add points to the plot, which gives us another hint towards a positive correlation due to the point cloud staying within a corridor parallel to the axes. Moreover, we can even see two clusters of wines that are separable on this attribute pair. Brushing these points shows that the clusters also has very distinct values on other attributes. They relate, e.g., to many *nonflavanoid phenols* and, for the most part, *low hue* and *OD280/OD315* values. The transition of the axis pair *flavanoids* and *total phenols* to a scatter plot confirms the assumption: two different positively correlated patterns become visible.

The second use case bases on a dataset of 345 observations of penguins on three islands [1]. It consists of only 4 continuous and 2 categorical attributes regarding different properties of these penguins. The dataset was chosen due to its well known and clear data patterns some of which can already be observed by just analyzing the polycurves of our implementation in Figure 5.10 (top). The top center parallel coordinates view between the attributes *flipper length* and *culmen depth* seems to show a typical crossing pattern indicating a negative correlation. But once our novel point projection is turned on (Figure 5.10 (bottom)) it becomes very clear that there are actually two clusters each showing a positive correlation – as can be quickly verified by transforming it smoothly into a scatter plot. The right star plot even reveals a third cluster and the left one shows the apparently strong positive correlation between *body mass* and *flipper length* in the upper left quadrant as already described in section 5.3 (both marked in purple in Figure 5.10).

The dataset of a combined cycle power plant's measurements and output energy (about 9 500 data, 5 attributes, see Figure 5.11)[71]. The *output energy* has been duplicated to better fit the requirements of our application and to offer quicker analysis between input and output variables. It shows a strong negative correlations between *energy output* and *temperature* and *energy output* and *exhaust vacuum*. Also, the corridor of points between *energy output* and *ambient pressure* hints at a positive correlation between these attributes. The lines also show a distinct cluster of low *exhaust vacuum* val-



(a) without scatter points



(b) with scatter points

Figure 5.10: Two screens of the penguin dataset: a) Without scatter points it seems to indicate a negative correlation between *flipper length* and *culmen depth* (purple cross). b) The points reveal that the pattern actually stems from two positively correlated clusters. The top right quadrant even exhibits a third cluster (purple ellispes).

ues. Highlighting the cluster reveals barely visible clusters in other attribute pairs, e.g., *ambient pressure* and *energy output*.

These cases show how the implementation of our concepts – the combination of straight parallel coordinates, polycurve star plot and point projection – supports in analyzing even obscured patterns. The ability to seamlessly transition to a scatter plot eliminates the additional analysis step and mental effort to switch context and getting oriented in a separate scatter plot for verifying found patterns.

5.8 User Study

We conducted a user study to evaluate how the draggable layout, the smooth transformation between parallel coordinates, star and scatter plots and the projected points support the users. 19 participants (5 female, 14 male) with basic to profound experience with visualization and a background in computer science took part. All participants successfully participated in an introductory visualization course and thus were familiar with the concepts of parallel coordinates, star plots and scatter plots.

The study consisted of 5 parts:(1) a short introduction to the visualization types, how correlation can be deduced from visual patterns, how the transformation and the point projection works and an introduction to the web application used for the study, (2) a test phase with 8 test questions to make sure the participant is familiar with correlation and cluster patterns in parallel coordinates and scatter plots, (3) a playground where the participant can play around with the application to gain confidence using it, (4) the main study consisting of 7 tasks that lead the participants through the analytic steps described in section 5.7, (5) an SUS questionnaire and particular questions about the usefulness of single features.

The main study consisted of 7 tasks with increasing difficulty. Three tasks required the users to estimate and find a certain type of correlation (see Figure 5.11 for an example). Another two tasks were about finding and selecting specific clusters. The remaining two tasks asked about complex multi-attribute relations. The easier tasks could be solved correctly by most participants (80-95%). For the more complex tasks the correct answers were in the range of 45-65%. Here participants identified cluster centers correctly in most cases but had issues with the specification of the outline of clusters which was somewhat cumbersome. Accordingly, some participants remarked that they would appreciate more fine-grained selection control to improve the precision of cluster selections.

The overall average SUS score for our system is 74.7 (SD = 13.5), which translates to a good usability. Particularly noteworthy are the good results regarding integration (question 5), consistency (question 6) and learnability (question 7) with average scores of 4.47 (SD = 0.6), 4.37 (SD = 0.9, inverted) and 4.26 (SD = 1.1). They show how much the integration of the scatter plot into the parallel coordinate system was appreciated and that the points were helpful without obscuring other information.

11 participants found the smooth transitions to be particularly useful. One of them especially appreciated the continuous control over the transition; another remarked that being able to transition to the scatter plot greatly improved his confidence in interpreting the PCP patterns. Also, the selection tools (10 participants), the point projection (2 participants) and the adjustable layout (2 participants) were mentioned as greatly supporting the given tasks. 9 participants missed the ability to switch axes which was not made available to simplify the interface. Additional features like a more

fine-grained selection control (3 participants), snapping the axes to a grid (2 participants) and flexible spacing of axes (1 participant) have been proposed after the study. Two participants reported that they found the technique fun to use and found themselves playing around with the data on their own accord. Another participant even felt "[...] like visual analytic experts would really appreciate this system", another agreed in a short oral recap after the study.

Another five questions surveyed the usefulness of specific parts of the design to be answered with a Likert scale from 1 to 5. Through them participants reported that dragging the axes was easy to do (M = 4.5, SD = 0.6) and useful (M = 4.8, SD = 0.5), transitioning from parallel coordinates to the scatter plot was simple to follow and understand (M = 4.5, SD = 0.5) and the scatter points supported the participants in figuring out clusters (M = 4.2, SD = 1.0) and correlations (M = 4.2, SD = 0.9).

Overall, the results from the tasks and the participants' reports regarding the application's support in correlation and cluster analysis show the great potential of the smooth integration of scatter plots and the point embedding to enhance parallel coordinate views.

5.9 Scalability and Appearance

The 9 500 data items of the power plant dataset [71] used in the user study (see Figure 5.11) push the technique to its limit in terms of overplotting (interaction still works well). The application's scalability depends essentially on its base visualizations' scalability. Thus, advanced techniques that enable scatter plots, star plots, and parallel coordinates to cope with a larger number of data items such as continuous parallel coordinates and scatter plots also improve the scalability of our approach.

The 13 attributes of the wine dataset [34] (see Figure 5.2) cap the number of attributes that could be analyzed at the same time. For a larger number of attributes, other configurations of rails – like a serpentine line or some off-screen space – might do a better job. The six attributes of the penguin dataset [1] (see Figure 5.10) are the recommended minimum for the closed rectangular form. Five attributes would work but reduce the number of simultaneously available orthogonal axis pairs to two but using an axis more than once is a sensible option in many cases.

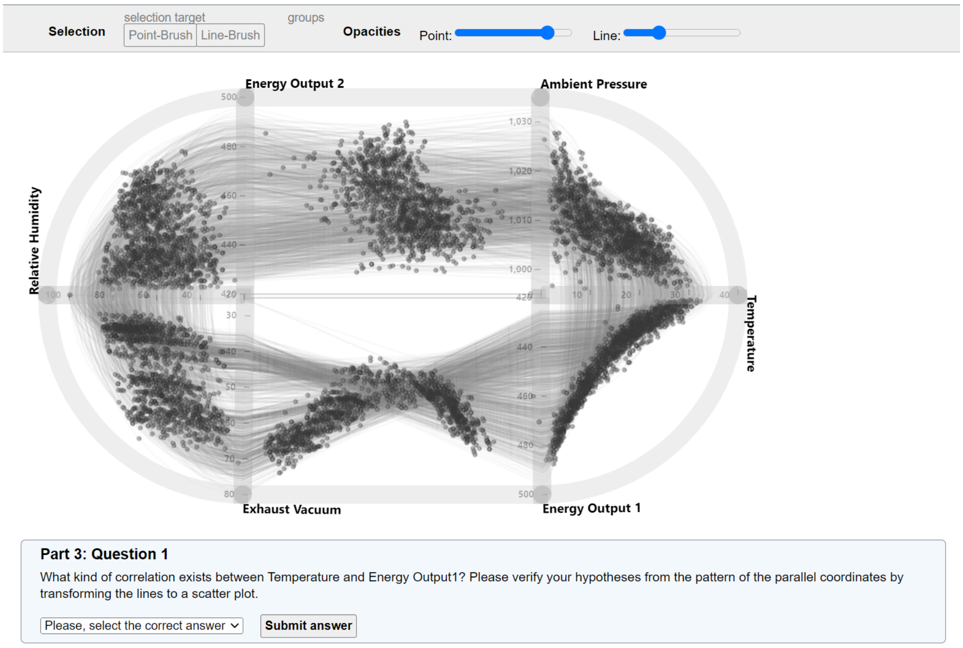


Figure 5.11: Web application with the main study's first question on the combined cycle power plant dataset [71]. The form of the point cloud between *temperature* and *energy output 1* indicates a strong negative correlation. We increased the width of the axis representations in comparison to the other depictions of the system to facilitate selecting and dragging the axes during the study.

5.10 Conclusion and Future Work

We presented seamless and geometrically grounded transitions between parallel coordinates and scatter plots via polycurve star plots. The polycurve star plot serves as a mediator visualization to allow for smooth transitions between parallel coordinates and scatter plots. It also improves the traceability of individual data items and cluster detection over regular star plots. The embedding of the scatter points facilitates the assessment of clusters and relations by adding valuable information to the usual line representation of parallel coordinate plots and star plots and provides another precise way to select data items from the data set. The smooth transitions in combination with the scatter points allow for visual tracking of various data configurations across plot types and therefore support the process of verifying clusters and correlations in the different plots. This was confirmed by the results of our user study, in particular the SUS score and also by the participants' comments especially emphasizing the smoothness of the transitions.

The flexible layout of the visualization based on rails allows for extending the technique in various ways. For example, some sections of the rails could be squeezed or expanded to provide a focus and context view. One could also imagine a four-sided layout with a larger rectangle in the middle which could host a control center or a semantically abstracted view of the data. While the added value of such enhancements depends on the actual dataset, the usability of the techniques needs to be evaluated in an application context where it is just one of many steps in the data analysis process.

Our next step addresses the scalability to larger datasets. To this end, we plan to investigate the applicability of our approach to cluster-based parallel coordinates and continuous parallel coordinates and scatter plots. However, even in its current form, our work contributes already to the portfolio of comprehensible and intuitive transitions between different visualization techniques.

5.11 Extension: Fine-grained Analysis of the AAE2 corpus

This section extends the original publication to underline the fitness of the visualization and interaction techniques to support the analysis of text corpora.

To gain a deeper insight into the interrelations of the argumentative essays in the AAE2 corpus [133], we extracted 8 statistical measures from each

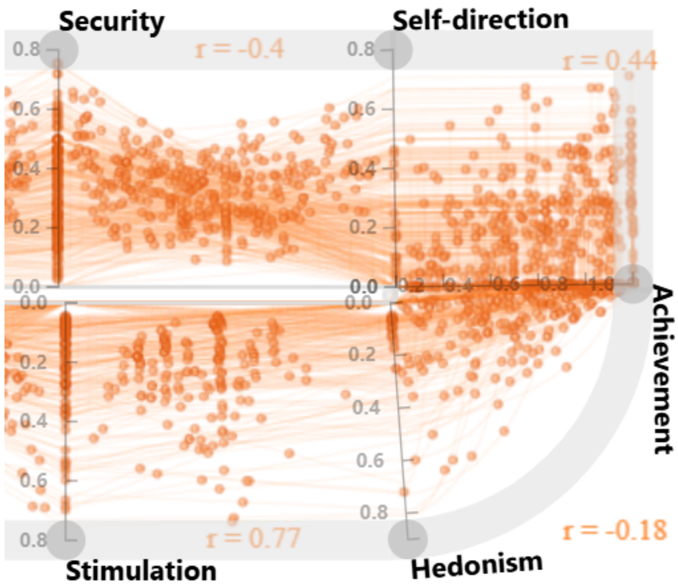
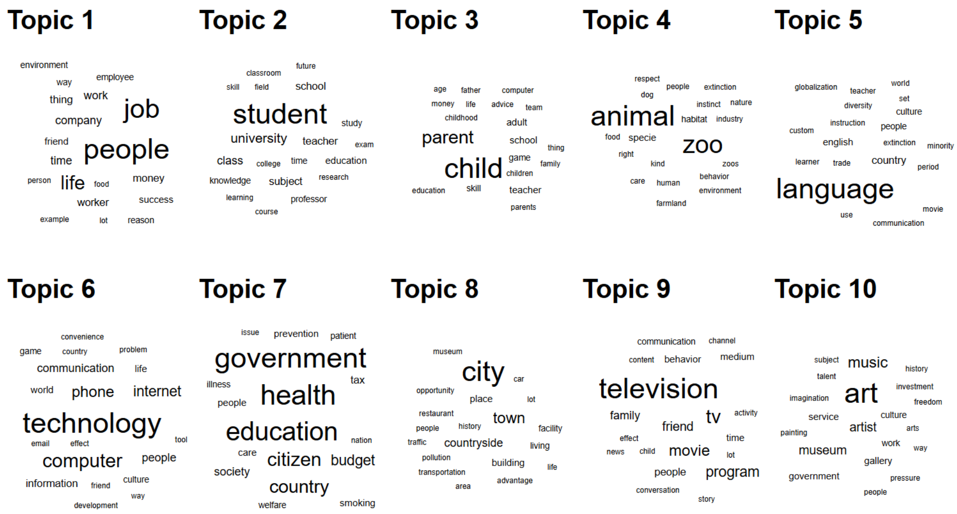


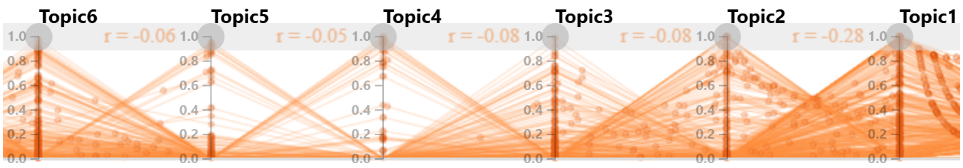
Figure 5.12: Analyzing the relations between five human values in the AAE2. The orange numbers denote the correlation coefficients between two adjacent human values. The lines and point clouds indicate a negative correlation between Security and Self-direction (the typical line crossings and the u-shape of the point cloud). A positive correlation can be seen between Hedonism and Stimulation (the point cloud partially forms a parallel line to the axes).

essay: number of pro and con units, premises, claims, and major claims, the total number of argument units, the maximum depth of the hierarchy, and the average number of child nodes in the hierarchy. To study the contents, we automatically extracted 10 topics using Non-Negative Matrix Factorization (NMF)[86] and 12 human values with the detector from Schroter et al. [23]. So in total, we have 30 attributes for all 402 essays of the AAE2 corpus. The dataset is uploaded to the analysis tool for an analysis of multi-attribute relations.

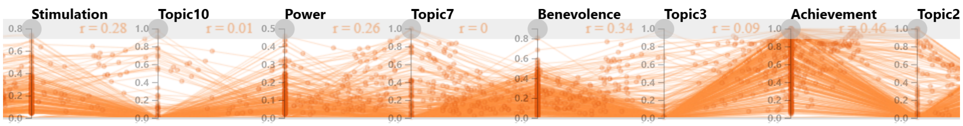
For the study of the relations between human values (from Kiesel et al. [65], Schwartz et al. [127]), the axes are rearranged such that the pairs of human values with the largest correlation (in absolute terms) are placed next to each other (see Figure 5.12 for an extract of the arrangement). The strongest correlation found is between Hedonism and Stimulation (r=0.77). Since Hedonism is about having pleasure and Stimulation is about having a varied and exciting life, it makes sense that one value supports the other. Similarly, Achievement (being successful and ambitious) and Self-direction (choosing one's own goals) tend to go together (r=0.44). Other combinations of human values are less likely to be found in one essay: Security and



(a) Word clouds of the 10 NMF-generated topics contained in the AAE2 corpus.



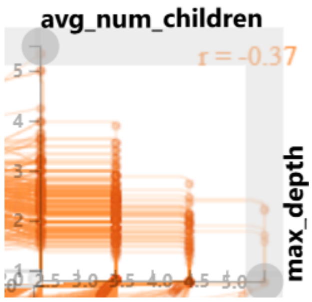
(b) The topics are for the most part not interrelated, as can be seen by the typical triangular shapes around each axis. The correlation coefficients (orange numbers) also show low values, except for the weak correlation between Topic 1 (work life) and 2 (school/university).



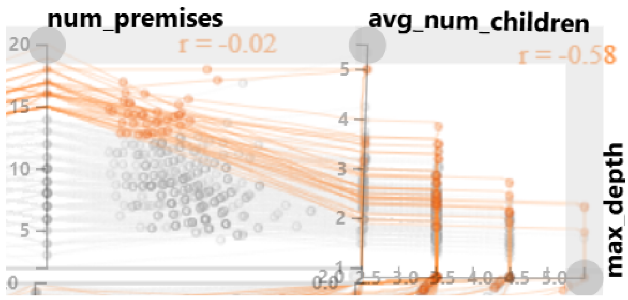
(c) Four pairs of human values and topics that are interrelated.

Figure 5.13: Analyzing the relations of NMF-generated topics in the AAE2.

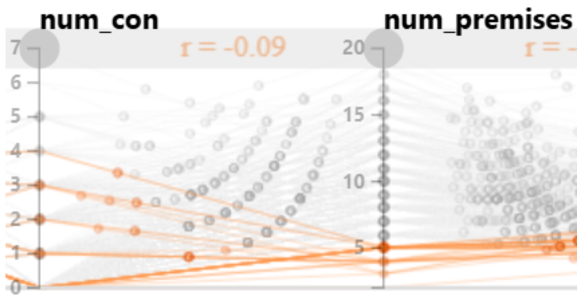
Self-direction show a negative correlation (r=-0.4), so the thought of a safe and stable home does not seem to go well with the idea of freedom and independence, at least in our essay corpus. The correlations partially reflect the compatibility of the human values found in Schwartz et al. [127]. Stimulation and Hedonism have a high compatibility and are, therefore, placed next to each other in Schwartz's system of basic human values. Self-direction and Achievement are also close to each other, while Security and Self-direction are maximally far apart. Achievement and Hedonism, however, even though close in Schwartz's system, show a slight negative correlation (r=-0.18) in our corpus. The different interpretation in our corpus might stem from the deviation of the data source: the original compatibility



(a) The relation between the average number of children and the maximum depth of the tree (r=-0.37).



(b) The correlation between the average number of children and argument depth gets more pronounced if only essays with 15 or more premises are selected.



(c) Essays with few premises can contain comparably many con units.

Figure 5.14: Analyzing the statistics of the argument units of the AAE2. As before, the orange numbers denote correlation coefficients.

between human values stems from survey questions about how well the human values go together, while our data is an automated extraction of human values from essays that might even contain several subtopics.

The analysis of topics yields only weak to no correlations (r$\leq$ $\pm$0.15), meaning that each essay is assigned to one main topic and varying minor topics. The visual result is the typical triangular shape around each axis that can be seen in Figure 5.13 (B). The only exceptions to this are Topics 1 (work life) and 2 (school/university), which show a weak negative correlation of r = -0.28. However, some interactions between topics and human values can be observed (Figure 5.13 (c)): Topic 2 (school/university) has a moderate positive correlation to the value Achievement (r=0.46), probably since many essays relate to being successful in education. Weaker relations can be found between Benevolence (be responsible and loyal) and Topic 3 (family life) (r=0.34), Power (have influence/wealth) and Topic 7 (government) (r=0.26), and Stimulation (have a varied life) and Topic 10 (art) (r=0.28). The analysis verifies, to some extent, that the argumentation in the essays draws on values that are typical of the topics they contain.

Within the simple statistics of argument units of each essay, we can verify and extend some of the insights gained in the last chapter. When looking at 5.14 (C), we find that essays with five or fewer premises still can contain high ratios of con units. The same conclusion can be drawn from the corpus overview visualization in Figure 2.4, where several short essays with many con units could be observed. The shape of the argumentation structure can be analyzed by investigating the scatter plot between the maximum depth and the average number of children (5.14 (A)). It shows a slight negative correlation between the attributes (r=-0.37). The strength of the correlation increases to -0.58 if only the essays with 15 or more premises are taken into account (see Figure 5.14 (B)). So, at least for longer essays, there seems to be two distinct strategies: either argue deeply (large maximum depth) or argue broadly (more children).

6
Conclusion

The presented research proposes solutions to common visualization problems regarding the exploratory analysis of argumentation and deliberation in text corpora.

6.1 Main Contributions

The main contributions of the thesis directly address the research questions outlined in the introduction.

How can argumentation structures be depicted so that users can compare argument strategies between (groups of) essays?

As established in the introduction, important components of an argument strategy, in addition to the argumentation structure, include the order and stance of argument units. The employed **tree aggregation technique** stacks the trees of argumentation structures such that the hierarchical positions and the order of the argument units are preserved. This is an important property, since it ensures that argument units with the same function are aggregated in the same node and, as such, allow meaningful analysis of the argument strategy.

The visual design of the nodes is crucial for the analysis of the set of argument units represented by the node, since it defines the analysis tasks that can be performed. For the analysis of argument units where the presence and stance need to be depicted, our research found **hierarchical histograms** to be the ideal choice. The stacked bars are easy to understand and can be estimated with high precision. Countables – small countable visual representatives of values that are too small to be reliably estimated from the bars – introduce a new way to increase the estimation precision for rare argument unit positions.

The **ArguLines** are compact pixel-based depictions of individual essays. The thesis developed them to abstract the text to a sequence of argument units, represented by their type, position, and stance, reducing the argumentation to its essential information. Their compact nature allows the analyst, for the first time, to get an overview of the essay corpus in a single view. In combination with filtering and ordering mechanisms, the technique simplifies complex comparisons of many different argumentation structures to visual clustering tasks.

How can the tree aggregation technique be extended to cope with the multitude of attributes per node studied in communication research?

In order to adapt the tree aggregation technique to the new data set and research tasks, both the way how to stack the tree and the design of the visual node representation were rethought. **Stacking threads** such that the order and positions of the comments within the thread structures are preserved summarizes comment sets of similar impact and function for the deliberative process, as confirmed by the domain experts.

The thesis developed a compound rose-chart-like glyph as **a suitable node representative**. The glyph is tailored to the dataset of our communication scientists and summarizes the properties of the represented comment set: the frequencies of up to 20 binary deliberative indicators hierarchically partitioned into 4 dimensions and the distribution of the ideological orientation. This way, communication scientists can analyze the relationship between deliberative quality indicators and ideological orientation. While specifically designed for this use case, the glyph can be used in other contexts, like the analysis of human values in essays in relation to the stance distribution of the essays.

The modular glyph design allows the manual configuration of which and how many indicators are shown, and can be adjusted to different sizes via **semantic zoom**. When the glyph is very small, for example when used as a node representative in the comparison of many tree aggregates, the glyph reduces itself to showing the distribution of ideological orientation, adding in aggregated information about the deliberative quality per dimension if the amount of space allows.

The full glyph design serves as the central component of the set shaping component that extends the tree aggregation by a versatile filter mechanism that allows the analyst to incrementally shape a set of relevant comments by subsequent splitting by selected attributes and merging the resulting subsets. The tree aggregation view reacts to the modifications by only showing threads that contain the selected set of comments.

Computational notebooks are a commonly used tool for computer scientists, but not yet established in the digital humanities. The requirements regarding the functionality of the notebooks differ due to a lack of programming skills; therefore, **a purely visual notebook** is a better fitting choice. Therefore, a visual notebook to store, analyze, compare, and comment on findings in thread sets has been developed and tested. The experts – both associated and independent – appreciated the ability to store findings in fully functional visualizations within the notebook, along with textual comments to document their research.

Does an area or a length encoding of the fill level of the circle sectors in rose charts align better with human perception?

The design of the rose-chart-like glyph prompted a research question concerning the perception mechanisms involved in estimating values from rose charts. The two studies revealed that neither pure area nor pure length estimation is used when estimating values from the fill level of circle sectors: It is **a mixture of both**. Below 65% fill level, area perception is dominant, while for larger values, there is a transition to length perception. Probably, when the distance between the outer border of the chart and the fill level gets smaller, the distance is taken into account when doing value estimations.

How can we effectively combine scatter plots and parallel coordinate plots to support the visual statistics evaluation of argumentation?

A seamless transition between parallel coordinates and scatter plots combines the strength of the scatter plot to show two-attribute relations and clusters with the ability of the parallel coordinates to analyze multi-attribute relations.

The thesis developed a new smooth transition **based on geometric observations**: (1) (Polycurve) star plots share the layout of axes with scatter plots and the way of representing data items with parallel coordinates which mark them as ideal mediator between the visualization types; (2) Cubic Bézier curves can smoothly transform a curved segment from a polycurve star plot to two orthogonal straight lines intersecting at the scatter point by shifting and increasing the weights of its control points. Since the transition is based on geometric observations, all in-between states are meaningful, and the transition itself can provide additional information, for example, regarding cluster memberships due to the Gestalt principle of common fate.

The transition naturally suggested a new way to enhance parallel coordinates with **scatter points**. Each scatter point projects the proportion of two values on two adjacent axes on the line connecting the values on those axes. The points blend in beautifully with the proposed transition and improve the capabilities of parallel coordinates to show clusters and correlations.

The dissertation also contributes a scientific examination of **polycurve star plots** that use polycurves instead of polylines to represent data items and their comparison to line-based star plots. The analysis revealed that the polycurves are better suited to express both positive and negative correlations as well as clusters. Additionally, the smoothness of the curve supports the traceability of individual data records through the plot.

6.1.1 Contributions to the Social Sciences

Since elaborate visualization, beyond line charts, bar charts, or scatter plots, is not commonly used in the social sciences, the visual and exploratory perspective inspired many experts who tested the developed visualization techniques. The **new way to see and interact with the data** also encouraged new ways to think and work with it. The visual approach advocates the exploration of the dataset and can reveal unusual or unexpected relations between data items that would not have been studied if the usual hypotheses-driven research approach had been followed. Many details about the change of deliberative indicators during a discussion or the pro/con patterns of the essays could not be perceived with such clarity when working with statistics and simple charts. A more comprehensive view of the data encourages interactive exploration and allows scientists to move beyond testing only preconceived, theory-driven hypotheses.

6.1.2 Contributions to Education

The expert reviews and user evaluations exploring the visualization and interaction techniques revealed their potential for serving educational purposes. The interplay between the aggregated argumentation structures and the individual argumentation visualizations can be used to explain the differences between high-rated and low-rated essays in class. Students can contrast their work with the best essays in the class to analyze where they could improve their writing as part of a self-assessment. The visual representation of the argument units improves understanding and also communication about the general structure of an argument and typical fallacies. It allows for contrasting different argumentative strategies or argumentation schemes (for example, those of Walton [150]).

The Comment Analyzer might support educators in teaching norm dynamics and deliberative processes in online debates. The richness of the glyph allows educators to show how different deliberative indicators behave in relation to each other, how the political orientation, norm groups, and the topic influence the quality of the deliberative process in a thread, and how the design of the online forum affects the structure of the threads. The students can become active researchers themselves and curate a data story based on their findings with the interactive, purely visual notebook.

The smooth and geometrically motivated transition between scatter plots and parallel coordinate plots enables students to familiarize themselves with the relationship between the two visualization types. Since each has its ideal use cases in data analysis, proficiency in both techniques needs to be acquired in the study of data analysis. The transition supports students in building a mental map of how different patterns in one visualization appear in the other visualization and in learning to recognize relations and clusters both in scatter plots and parallel coordinate plots.

6.2 Text visualization and LLMs

During the development of the dissertation, new methods for natural language processing emerged. Large language models improved the automatic generation, summarization, and understanding of text drastically. Even visualizations can be generated with multimodal models. One might ask oneself whether developing visual analytics systems for text still makes sense, if a question to an LLM suffices to get the same information, just quicker, better formulated, and accompanied by a suitable visualization.

The worth of a visual analysis system lies, as stated before, in its ability to allow for an exploration of the data. When asking an LLM, some preconception about the data needs to be present to formulate the question. If the research question is not clear yet, a visual overview of the data supports generating new and relevant ones in a data-driven way. The visual representations that an LLM can generate for depicting derived values are limited to variations of known representations, since it can only draw from information it has been trained on. Tailored designs and novel visualization techniques still need human expertise.

LLMs can increase the worth of the proposed visual analytics techniques by supporting annotators in extracting the structures and text attributes from the text. A combination of training examples and instructions for the definition of each type of argument unit, human value, topic, and deliberative indicator might be enough to teach an LLM how to find this information in unknown texts. With an LLM-based automation of the annotation process, the analysis of common or unusual phenomena in argumentation could easily be extended to argumentation of different nationalities and languages or a multitude of different topics.

So, while LLMs can solve many textual and simple visualization tasks, they will not be able to substitute the research domain of visual text analytics, but might help in extending the currently still rather rare argument-annotated corpora for a broader analysis of argumentation structures.

6.3 Open Problems and Future Work

The thesis has raised new research questions that could not be covered in the dissertation: The visually meaningful aggregation of tree structures has been a core concept in the thesis. The dissertation focused on the expressiveness of the nodes to summarize the key attributes of the represented set of comments or argument units. But, the links might also offer potential to add additional information, like the time that elapsed between the posts of connected comments in an online conversation or the type of relation between argument units. Also, a way to unambiguously identify individual trees within the aggregation would visually support the analysis of groups of similar structure. For the hierarchical bar charts, this could be done by replacing the aggregated information of the bar with a pixel-based representation of each argumentation in the form of a barcode, where each argumentation is identified as one stripe. Ordering the barcode, e.g., by the stances of the argument units in a node can yield important information about the influence of the stance in the argument units of the respective node on the properties / property distributions of the argument units in other nodes. However, this would only work for a limited number of argumentation structures and one attribute, for example, the stance of the argument unit. New ideas are needed for more than one attribute.

Written arguments, especially in the political sciences, evolve. To illustrate which arguments are taken up by other authors and how they are changed in the process, new visualization techniques for capturing the migration of arguments over a corpus of publications need to be developed. An automated analysis of migration paths yields a graph that connects similar text passages. Each text passage might have multiple possible source text passages from which content might have migrated into the text passage and multiple target text passages that might have taken ideas or concepts from the text passage. A visualization that captures the distribution of those source and target text passages over an individual publication can reveal the research impact of different parts of a publication. To develop visualization and interaction techniques to trace arguments from document to document, to zoom out to see an overview of possibly migrating arguments along a timeline, and zoom in again to focus on a specific document or argument is still future work.

Humankind will never stop arguing. Many of the complex processes and language constructs of argumentation are still subject to active research in social sciences and linguistics. Some of the complexity could be captured, made tangible, and approachable using the techniques developed and evaluated in this dissertation.

Bibliography

[1] Allison M. Horst, Alison P. Hill, and Kristen B. Gorman. palmerpenguins: Palmer Archipelago (Antarctica) penguin data, 2020. URL https://allisonhorst.github.io/palmerpenguins/.

[2] amCharts. Radial Histogram, 2024. https://www.amcharts.com/demos-v4/radial-histogram-v4/ [Accessed: 2024-01-16].

[3] Pablo Aragón, Vicenç Gómez, and Andreas Kaltenbrunner. Detecting Platform Effects in Online Discussions. *Policy & Internet*, 9(4): 420–443, 2017. doi: 10.1002/poi3.158.

[4] Aristotle. *On Rethoric, A Theory of Civic Discourse*. Clarendon Aristotle series, 2007. doi: 10.1080/07350190701419913.

[5] Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49, 2015. doi: 10.1080/19462166.2014.1001791.

[6] Lyn Bartram and Maureen C. Stone. Whisper, Don't Scream: Grids and Transparency. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1444–1458, 2011. doi: 10.1109/TVCG.2010.237.

[7] Tony Bergstrom and Karrie Karahalios. Conversation Clusters: Grouping Conversation Topics Through Human-computer Dialog. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2349–2352, New York, NY, USA, 2009. doi: 10.1145/1518701.1519060.

[8] Anastasia Bezerianos, Fanny Chevalier, Pierre Dragicevic, Niklas Elmqvist, and Jean-Daniel Fekete. Graphdice: A system for exploring multivariate social networks. In *Computer graphics forum*, volume 29, pages 863–872, 2010. doi: 10.1111/j.1467-8659.2009.01687.x.

[9] Bongshin Lee, Cynthia S. Parr, Catherine Plaisant, Benjamin B. Bederson, Vladislav D. Veksler, Wayne D. Gray, and Christopher Kotfila. TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts. *IEEE Transactions on Visualization and Computer Graphics*, 12 (6):1414–1426, 2006. doi: 10.1109/TVCG.2006.106.

[10] Rita Borgo, Johannes Kehrer, David H. S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics 2013 - State of the Art Reports*, pages 39–63. The Eurographics Association, 2013. doi: 10.2312/conf/EG2013/stars/039-063.

[11] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013. doi: 10.1109/TVCG.2013.234.

[12] Sebastian Bremm, Tatiana von Landesberger, Martin Heß, Tobias Schreck, Philipp Weil, and Kay Hamacherk. Interactive visual comparison of multiple trees. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 31–40, 2011. doi: 10.1109/VAST.2011.6102439.

[13] John Brosz, Miguel A. Nacenta, Richard Pusch, Sheelagh Carpendale, and Christophe Hurter. Transmogrification: Causal Manipulation of Visualizations. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 97–106, New York, NY, USA, 2013. Association for Computing Machinery. doi: 10.1145/2501988.2502046.

[14] Teodorico Caporaso, Stanislao Grazioso, Giuseppe Di Gironimo, and Antonio Lanzotti. Biomechanical indices represented on radar chart for assessment of performance and infringements in elite racewalkers. *Sports Engineering*, 23(1):1–8, 2020. doi: 10.1007/s12283-019-0317-2.

[15] Bram C. M. Cappers and Jarke J. van Wijk. Exploring multivariate event sequences using rules, aggregations, and selections. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):532–541, 2017. doi: 10.1109/TVCG.2017.2745278.

[16] Isaac Cho, Ryan Wesslen, Svitlana Volkova, William Ribarsky, and Wenwen Dou. CrystalBall: A Visual Analytic System for Future Event Discovery and Analysis from Social Media Data. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 25–35. IEEE, 2017. doi: 10.1109/VAST.2017.8585658.

[17] Christian Tominski, Gennady Andrienko, Natalia Andrienko, Susanne Bleisch, Sara Irina Fabrikant, Eva Mayr, Silvia Miksch, Margit Pohl, and André Skupin. Toward flexible visual analytics augmented through smooth display transitions. *Visual Informatics*, 5(3):28–38, 2021. doi: 10.1016/j.visinf.2021.06.004.

[18] Christian Voigt, Gregor Betz, Helen Bohse, Sebastian Cacean, and David Schneider. Argunet, 2018. URL `http://www.argunet.org/`.

[19] Jarry H.T. Claessen and Jarke J. van Wijk. Flexible Linked Axes for Multivariate Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2310–2316, 2011. doi: 10.1109/TVCG.2011.201.

[20] William S. Cleveland and Robert McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: 10.2307/2288400.

[21] Christopher Collins, Antske Fokkens, Andreas Kerren, Chris Weaver, and Angelos Chatzimparmpas. Visual Text Analytics (Dagstuhl Seminar 22191). *Dagstuhl Reports*, 12(5):37–91, 2022. doi: 10.4230/DagRep.12.5.37. URL `https://drops.dagstuhl.de/opus/volltexte/2022/17443`.

[22] Jeff Conklin and Michael L. Begeman. gIBIS: A Hypertext Tool for Exploratory Policy Discussion. *ACM Transactions on Information Systems*, 6(4):303–331, 1988. doi: 10.1145/58566.59297.

[23] Daniel Schroter, Daryna Dementieva, and Georg Groh. Adam-Smith at SemEval-2023 Task 4: Discovering Human Values in Arguments with Ensembles of Transformer-based Models. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval'23)*, pages 532–541. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.semeval-1.74.

[24] Yu Dong, Alex Fauth, Maolin Huang, Yi Chen, and Jie Liang. PansyTree: Merging Multiple Hierarchies. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pages 131–135. IEEE, 2020. doi: 10.1109/PacificVis48177.2020.1007.

[25] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. The Information Flaneur: A Fresh Look at Information Seeking. In *Proc. of*

the SIGCHI Conference on Human Factors in Computing Systems, pages 1215–1224. ACM, 2011. doi: 10.1145/1978942.1979124.

[26] Jose Díaz, Oscar Meruvia-Pastor, and Pere-Pau Vázquez. Improving Perception Accuracy in Bar Charts with Internal Contrast and Framing Enhancements. In *2018 22nd International Conference Information Visualisation (IV)*, pages 159–168. IEEE, 2018. doi: 10.1109/iV.2018.00037.

[27] Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel Keim. ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum*, 35 (3):431–440, 2016. doi: 10.1111/cgf.12919.

[28] Mennatallah El-Assady, Rita Sevastjanova, Bela Gipp, Daniel Keim, and Christopher Collins. NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations. *Computer Graphics Forum*, 36 (3):213–225, 2017. doi: 10.1111/cgf.13181.

[29] Mennatallah El-Assady, Rita Sevastjanova, Daniel Keim, and Christopher Collins. ThreadReconstructor: Modeling Reply-Chains to Untangle Conversational Text through Visual Analytics. *Computer Graphics Forum*, 37(3):351–365, 2018. doi: 10.1111/cgf.13425.

[30] E. Fanea, S. Carpendale, and T. Isenberg. An interactive 3D integration of parallel coordinates and star glyphs. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 149–156, 2005. doi: 10.1109/INFVIS.2005.1532141.

[31] Federica Di Castro and Enrico Bertini. Surrogate Decision Tree Visualization. In *Joint Proc. of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces (ACM IUI 2019), Los Angeles, USA, March 20, 2019*, pages 1–5. CEUR-WS.org, 2019.

[32] Martin Fink, Jan-Henrik Haunert, Joachim Spoerhase, and Alexander Wolff. Selecting the Aspect Ratio of a Scatter Plot Based on Its Delaunay Triangulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2326–2335, 2013. doi: 10.1109/TVCG.2013.187.

[33] Danyel Fisher. Animation for Visualization: Opportunities and Drawbacks. In *Beautiful Visualization*. O'Reilly Media, 2010. ISBN 978-9350230510.

[34] Michele Forina, Riccardo Leardi, Armanino C, and Sergio Lanteri. *Parvus: An Extendable Package Of Programs For Data Exploration, Classification And Correlation*. Elsevier Science, 1998. ISBN 0-444-43012-1.

[35] Frank van Ham, Martin Wattenberg, and Fernanda B.Viégas. Mapping Text with Phrase Nets. *IEEE Transactions on Visualization and Computer Graphics*, 15(06):1169–1176, 2009. doi: 10.1109/TVCG.2009.165.

[36] James B. Freeman. An Approach to Argument Macrostructure. In *Argument Structure*, volume 18 of *Argumentation Library*, pages 1–38. Springer Netherlands, Dordrecht, 2011. ISBN 978-94-007-0356-8. doi: 10.1007/978-94-007-0357-5.

[37] Siwei Fu, Jian Zhao, Weiwei Cui, and Huamin Qu. Visual Analysis of MOOC Forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):201–210, 2017. doi: 10.1109/TVCG.2016.2598444.

[38] Siwei Fu, Yong Wang, Yi Yang, Qingqing Bi, Fangzhou Guo, and Huamin Qu. VisForum: A Visual Analysis System for Exploring User Groups in Online Forums. *ACM Trans. Interact. Intell. Syst.*, 8(1), 2018. doi: 10.1145/3162075.

[39] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 3237–3246. ACM, 2013. doi: 10.1145/2470654.2466443.

[40] Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, Fabian Fischer, and Enrico Bertini. The influence of contour on similarity perception of star glyphs. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2251–2260, 2014. doi: 10.1109/TVCG.2014.2346426.

[41] Michael Glueck, Mahdi Pakdaman Naeini, Finale Doshi-Velez, Fanny Chevalier, Azam Khan, Daniel Wigdor, and Michael Brudno. PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):371–381, 2018. doi: 10.1109/TVCG.2017.2745118.

[42] Valentin Gold, Christian Rohrdantz, and Mennatallah El-Assady. Exploratory Text Analysis using Lexical Episode Plots. In *Eurograph-*

ics Conference on Visualization (EuroVis), pages 85–89, 2015. doi: 10.2312/eurovisshort.20151130.

[43] Jean Goodwin. Wigmore's Chart Method. *Informal Logic*, 20(3), 2000. ISSN 0824-2577. doi: 10.22329/il.v20i3.2278.

[44] Martin Graham and Jessie Kennedy. Using Curves to Enhance Parallel Coordinate Visualisations. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003*, volume 2003, pages 10–16, 2003. doi: 10.1109/IV.2003.1217950.

[45] Martin Graham and Jessie Kennedy. A survey of multiple tree visualisation. *Information Visualization*, 9(4):235–252, 2010. doi: 10.1057/ivs.2009.29.

[46] Henning Gruendl, Patrick Riehmann, Yves Pausch, and Bernd Froehlich. Time-Series Plots Integrated in Parallel-Coordinates Displays. *Computer Graphics Forum*, 35(3):321–330, 2016. ISSN 0167-7055. doi: 10.1111/cgf.12908.

[47] Ivan Habernal and Iryna Gurevych. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1):125–179, 2017. doi: 10.1162/COLIa00276.

[48] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. Ranking Visualizations of Correlation Using Weber's Law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1943–1952, 2014. doi: 10.1109/TVCG.2014.2346979.

[49] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 203–212. ACM, 2010. doi: 10.1145/1753326.1753357.

[50] Jeffrey Heer and George Robertson. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007. doi: 10.1109/TVCG.2007.70539.

[51] Julian Heinrich and Daniel Weiskopf. State of the Art of Parallel Coordinates. In *Eurographics (STARs)*, pages 95–116, 2013. doi: 10.2312/conf/EG2013/stars/095-116.

[52] Henning Wachsmuth. *Text Analysis Pipelines—Towards Ad-hoc Large-scale Text Mining*, volume 9383 of *Lecture Notes in Computer Science*. Springer, 2015. doi: 10.1007/978-3-319-25741-9.

[53] Henning Wachsmuth, Martin Trenkmann, Benno Stein, and Gregor Engels. Modeling Review Argumentation for Robust Sentiment Analysis. In *Proc. of the 25th International Conference on Computational Linguistics (COLING 2014)*, pages 553–564, 2014.

[54] Henning Wachsmuth, Johannes Kiesel, and Benno Stein. Sentiment Flow—A General Model of Web Review Argumentation. In *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pages 601–611, 2015. doi: 10.18653/v1/D15-1072.

[55] Henning Wachsmuth, Giovanni Da San Martino, Dora Kiesel, and Benno Stein. The Impact of Modeling Overall Argumentation with Tree Kernels. In *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 17)*, pages 2369–2379, 2017. doi: 10.18653/v1/D17-1253.

[56] Nathalie Henry, Jean-Daniel Fekete, and Michael J. McGuffin. NodeTrix: a Hybrid Visualization of Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, 2007. doi: 10.1109/TVCG.2007.70582.

[57] Danny Holten and Jarke J. van Wijk. Evaluation of Cluster Identification Performance for Different PCP Variants. *Computer Graphics Forum*, 29(3):793–802, 2010. ISSN 0167-7055. doi: 10.1111/j.1467-8659.2009.01666.x.

[58] Enamul Hoque and Giuseppe Carenini. ConVis: A Visual Text Analytic System for Exploring Blog Conversations. *Computer Graphics Forum*, 2014. doi: 10.1111/cgf.12378.

[59] Enamul Hoque and Giuseppe Carenini. ConVisIT: Interactive Topic Modeling for Exploring Asynchronous Online Conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 169–180. ACM, 2015. doi: 10.1145/2678025.2701370.

[60] Enamul Hoque and Giuseppe Carenini. MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations. In *Proc. of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pages 96–107, New York, NY, USA, 2016. doi: 10.1145/2856767.2856782.

[61] Houssem B. Lahmar and Melanie Herschel. Structural summaries for visual provenance analysis. In *11th International Workshop on Theory*

and Practice of Provenance (TaPP 2019), page 2, Philadelphia, PA, 2019. USENIX Association. URL https://www.usenix.org/conference/tapp2019/presentation/lahmar.

[62] Ines Engelmann, Hanna Marzinkowski, and Klara Langmann. Salient deliberative norm types in comment sections on news sites. *New Media & Society*, 26(2):921–940, 2024. doi: 10.1177/14614448211068104.

[63] Alfred Inselberg and Bernard Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the First IEEE Conference on Visualization: Visualization90*, pages 361–378, 1990.

[64] Wolfgang Jentner, Mennatallah El-Assady, Bela Gipp, and Daniel A. Keim. Feature Alignment for the Analysis of Verbatim Text Transcripts. *EuroVis Workshop on Visual Analytics (EuroVA)*, pages 13–17, 2017. doi: 10.2312/eurova.20171113.

[65] Johannes Kiesel, Milad Alshomary, Nicolas Handke, Xiaoni Cai, Henning Wachsmuth, and Benno Stein. Identifying the Human Values behind Arguments. In *60th Annual Meeting of the Association for Computational Linguistics (ACL 2022)*, pages 4459–4471. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.acl-long.306.

[66] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, volume 1, pages 117–121, 2015. doi: 10.1109/PACIFICVIS.2015.7156366.

[67] Linda T. Kaastra and Brian Fisher. Field Experiment Methodology for Pair Analytics. In *Proc. of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 152–159. ACM, 2014. doi: 10.1145/2669557.2669572.

[68] Kang Xie. and Bijaya Karki. A Hybrid Approach based on Parallel Coordinates and Star Plot. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - IVAPP*, pages 267–274. SciTePress, 2019. doi: 10.5220/0007375502670274.

[69] Areti Karamanou, Nikolaos Loutas, and Konstantinos Tarabanis. ArgVis: structuring political deliberations using innovative visuali-

sation technologies. *Electronic Participation*, pages 87–98, 2011. doi: 10.1007/978-3-642-23333-38.

[70] Edna L. Kaufman, Miles W. Lord, Thomas Whelan Reese, and John Volkmann. The discrimination of visual number. *The American journal of psychology*, 62(4):498–525, 1949. doi: 10.2307/1418556.

[71] Heysem Kaya, Pmar Tüfekci, and Fikret S. Gürgen. Local and global learning methods for predicting power of a combined gas & steam turbine. In *Proceedings of the international conference on emerging trends in computer and electronics engineering ICETCEE*, pages 13–18, 2012.

[72] Mandy Keck, Dietrich Kammer, Thomas Gründer, Thomas Thom, Martin Kleinsteuber, Alexander Maasch, and Rainer Groh. Towards Glyph-Based Visualizations for Big Data Clustering. In *Proc. of the 10th International Symposium on Visual Information Communication and Interaction*, pages 129–136. Association for Computing Machinery, 2017. doi: 10.1145/3105971.3105979.

[73] Dora Kiesel. Video on the main concepts and workflows of the Comment Analyser, 2024. URL `https://www.uni-weimar.de/fileadmin/user/fak/medien/professuren/Virtual_Reality/videos/publications/CommentAnalyzerV6small.mp4`.

[74] Dora Kiesel, Patrick Riehmann, Henning Wachsmuth, Benno Stein, and Bernd Froehlich. Visual Analysis of Argumentation in Essays. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1139–1148, 2021. doi: 10.1109/TVCG.2020.3030425.

[75] Dora Kiesel, Patrick Riehmann, and Bernd Froehlich. Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots. *Computer Graphics Forum*, 42(6):e14923, 2023. doi: https://doi.org/10.1111/cgf.14923. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14923`.

[76] Dora Kiesel, Patrick Riehmann, Ines Engelmann, Hanna Ramezani, and Bernd Froehlich. Comment Analyzer: A Tool for Analyzing Comment Sets and Thread Structures of News Articles. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–13, 2025. doi: 10.1109/TVCG.2025.3544733.

[77] Dora Kiesel, Patrick Riehmann, and Bernd Froehlich. Rose Charts: Area or Length Encoding for Fill Level of Circle Sectors? In *EuroVis*

2025 - Short Papers. The Eurographics Association, 2025. doi: 10.2312/evs.20251075.

[78] Younghoon Kim and Jeffrey Heer. Gemini: A Grammar and Recommender System for Animated Transitions in Statistical Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):485–494, 2021. doi: 10.1109/TVCG.2020.3030360.

[79] Klaus Bayer. *Argument und Argumentation. Logische Grundlagen der Argumentationsanalyse (Studienbücher zur Linguistik, Band 1)*. Vandenhoeck & Ruprecht, 2 edition, 2007.

[80] Kenan Koc, Andrew Stephen McGough, and Sara Johansson Fernstad. PeaGlyph: Glyph design for investigation of balanced data structures. *Information Visualization*, 21(1):74–92, 2022. doi: 10.1177/14738716211050602.

[81] Robert Kosara. Evidence for Area as the Primary Visual Cue in Pie Charts. In *2019 IEEE Visualization Conference (VIS)*, pages 101–105, 2019. doi: 10.1109/VISUAL.2019.8933547.

[82] Robert Kosara. The impact of distribution and chart type on part-to-whole comparisons. In *Short Paper Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization (EuroVis)*, 2019. doi: 10.2312/evs20191162.

[83] Robert Kosara. Notebooks for data analysis and visualization: Moving beyond the data. *IEEE Computer Graphics and Applications*, 43(1): 91–96, 2023.

[84] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. DoSVis: Document Stance Visualization. In *Proc. of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: IVAPP*, pages 168–175, 2018. doi: 10.5220/0006539101680175.

[85] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. The State of the Art in Sentiment Visualization. *Computer Graphics Forum*, 37(1): 71–96, 2018. ISSN 0167-7055. doi: 10.1111/cgf.13217.

[86] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Proceedings of the 14th International Conference on Neural Information Processing Systems*, NIPS'00, page 535–541, Cambridge, MA, USA, 2000. MIT Press. doi: 10.5555/3008751.3008829.

[87] Dirk J. Lehmann and Holger Theisel. Orthographic Star Coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2615–2624, 2013. doi: 10.1109/TVCG.2013.182.

[88] Gilly Leshed, Jeffrey T. Hancock, Dan Cosley, Poppy L. McLeod, and Geri Gay. Feedback for Guiding Reflection on Teamwork Practices. In *Proc. of the 2007 International ACM Conference on Supporting Group Work*, GROUP '07, pages 217–220, 2007. doi: 10.1145/1316624.1316655.

[89] Zipeng Liu, Shing Hei Zhan, and Tamara Munzner. Aggregated Dendrograms for Visual Comparison between Many Phylogenetic Trees. *IEEE Transactions on Visualization and Computer Graphics*, 26(9):2732–2747, 2020. doi: 10.1109/TVCG.2019.2898186.

[90] John D. Lowrance, Ian W. Harrison, and Andres C. Rodriguez. Structured argumentation for analysis. In *Proc. of the 12th International Conference on Systems Research, Informatics, and Cybernetics*, pages 47–57, 2000. doi: 10.1080/19462166.2013.869764.

[91] Thoughtgraph Ltd. DebateGraph, Accessed 2018. URL `https://debategraph.org`.

[92] Timothy Luciani, Andrew Burks, Cassiano Sugiyama, Jonathan Komperda, and G. Elisabeta Marai. Details-First, Show Context, Overview Last: Supporting Exploration of Viscous Fingers in Large-Scale Ensemble Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1225–1235, 2019. doi: 10.1109/tvcg.2018.2864849.

[93] Martin Wattenberg and Fernanda B. Viégas. The Word Tree, an Interactive Visual Concordance. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1221–1228, 2008. doi: 10.1109/TVCG.2008.172.

[94] Rafael Messias Martins, Vasiliki Simaki, Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. StanceXplore: Visualization for the Interactive Exploration of Stance in Social Media. In *Proc. of 2nd Workshop on Visualization for the Digital Humanities*, pages 1–5. IEEE, 2017.

[95] Hanna Marzinkowski and Ines Engelmann. Die Wirkung "guter" Argumente. *Publizistik*, 63(2):269–287, 2018. doi: 10.1007/s11616-018-0408-2.

[96] Hanna Marzinkowski and Ines Engelmann. Rational-critical user discussions: how argument strength and the conditions set by news organizations are linked to (reasoned) disagreement. *Digital Journalism*, 10(3):433–451, 2022. doi: 10.1080/21670811.2021.1957968.

[97] Massimo Stafoggia, Adele Lallo, Danilo Fusco, Anna P. Barone, Mariangela D'Ovidio, Chiara Sorge, and Carlo A. Perucci. Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. *Journal of clinical epidemiology*, 64(7):770–778, 2011. ISSN 0895-4356. doi: 10.1016/j.jclinepi.2010.10.009.

[98] J. Matute and L. Linsen. Hinted Star Coordinates for Mixed Data. *Computer Graphics Forum*, 39(1):117–133, 2020. ISSN 0167-7055. doi: 10.1111/cgf.13666.

[99] Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2227–2236, 2013. doi: 10.1109/TVCG.2013.200.

[100] Tamara Munzner, François Guimbretière, Serdar Tasiran, Li Zhang, and Yunhong Zhou. TreeJuxtaposer: scalable tree comparison using Focus+ Context with guaranteed visibility. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 453–462, 2003. doi: 10.1145/1201775.882291.

[101] Florence Nightingale. Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army, 1858. URL `https://www.rct.uk/collection/1075240/notes-on-matters-affecting-the-health-efficiency-and-hospital-administration-of`. [Digital Version, Accessed: 2024-01-16].

[102] Niklas Elmqvist, Andrew Vande Moere, Hans-Christian Jetter, Daniel Cernea, Harald Reiterer, and TJ Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, 2011. doi: 10.1177/1473871611413180.

[103] Hui Niu. *Pedagogical efficacy of argument visualization tools*. PhD thesis, Simon Fraser University, Faculty of Education, 2016.

[104] Michael O'Donnell. RSTTool 2.4: a markup tool for Rhetorical Structure Theory. In *Proc. of the 6th European Workshop on Natural Language Generation*, pages 253–256, 2000. ISBN 965-90296-0-8. doi: 10.3115/1118253.1118290.

[105] Tomasz Opach, Stanislav Popelka, Jitka Dolezalova, and Jan Ketil Rød. Star and polyline glyphs in a grid plot and on a map display: which perform better? *Cartography and Geographic Information Science*, 45(5):400–419, 2018. doi: 10.1080/15230406.2017.1364169.

[106] OriginLab Corporation. Origin Help: Radial Bar and Radial Stacked Bar Plot, 2024. `https://www.originlab.com/doc/Origin-Help/Radial-Bar-Plot` [Accessed: 2024-01-16].

[107] Jack Paulus. Truthmapping, Accessed 2020. URL `https://www.truthmapping.com/`.

[108] Andreas Peldszus and Manfred Stede. An annotated corpus of argumentative microtexts. In *Proc. of the 1st European Conference on Argumentation*. College Publications, 2016.

[109] Lisa Perkhofer and Peter Hofer. Optimizing Scatterplot-Matrices for Decision-Support. In Fred D. Davis, René Riedl, Jan vom Brocke, Pierre-Majorique Léger, Adriane B. Randolph, and Gernot Müller-Putz, editors, *Information Systems and Neuroscience*, pages 63–76, Cham, 2021. Springer International Publishing. ISBN 978-3-030-88900-5.

[110] Scott Plous. *The psychology of judgment and decision making*. Mcgraw-Hill Book Company, 1993.

[111] John L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133(1-2):233–282, 2001. ISSN 00043702. doi: 10.1016/s0004-3702(01)00145-x.

[112] Jörg Pührer. ArgueApply: A Mobile App for Argumentation. In *Proc. of 14th International Conference on Logic Programming and Nonmonotonic Reasoning*, pages 250–262, 2017. doi: 10.1007/978-3-319-61660-523.

[113] Puripant Ruchikachorn and K. Mueller. Learning Visualizations by Analogy: Promoting Visual Literacy through Visualization Morphing. *IEEE Transactions on Visualization and Computer Graphics*, 21:1028–1044, 2015. doi: 10.1109/TVCG.2015.2413786.

[114] Raidou Renata Georgia, Eisemann Martin, Breeuwer Marcel, Eisemann Elmar, and Vilanova Anna. Orientation-Enhanced Parallel Coordinate Plots. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):589–598, 2016. doi: 10.1109/TVCG.2015.2467872.

[115] Chris Reed and Derek Long. Content ordering in the generation of persuasive discourse. In *Proceedings of the Fifteenth International Joint Conference on Artifical Intelligence - Volume 2*, IJCAI'97, page 1022–1027, San Francisco, CA, USA, 1997. Morgan Kaufmann Publishers Inc. doi: 10.5555/1622270.1622304.

[116] Chris Reed and Glenn Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979, 2004. doi: 10.1142/S0218213004001922.

[117] Severino Ribecca. The Data Visualisation Catalogue - Radial Column Chart, 2024. https://datavizcatalogue.com/methods/radial_column_chart.html [Accessed: 2024-01-16].

[118] Patrick Riehmann, Henning Gruendl, Bernd Froehlich, Martin Potthast, Martin Trenkmann, and Benno Stein. The NETSPEAK WORDGRAPH: Visualizing keywords in context. In *2011 IEEE Pacific Visualization Symposium*, pages 123–130, 2011. doi: 10.1109/PACIFICVIS.2011.5742381.

[119] Patrick Riehmann, Henning Gruendl, Martin Potthast, Martin Trenkmann, Benno Stein, and Bernd Froehlich. WORDGRAPH: Keyword-in-Context Visualization for NETSPEAK's Wildcard Search. *IEEE Transactions on Visualization and Computer Graphics*, 18 (9):1411–1423, 2012. doi: 10.1109/TVCG.2012.96.

[120] Patrick Riehmann, Jens Opolka, and Bernd Froehlich. The product explorer: decision making with ease. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 423–432, 2012. doi: 10.1145/2254556.2254638.

[121] Patrick Riehmann, Martin Potthast, Benno Stein, and Bernd Froehlich. Visual assessment of alleged plagiarism cases. *Computer Graphics Forum*, 34(3):61–70, 2015. doi: 10.1111/cgf.12618.

[122] Robert G. Lawson. Order of Presentation as a Factor in Jury Persuasion. *Kentucky Law Journal*, (56):523–555, 1968. URL https://uknowledge.uky.edu/law_facpub/239/.

[123] Ryan K Rowland. Radial Stacked Bar Charts in Tableau, 2016. https://ryankrowland.wordpress.com/2016/10/20/radial-stacked-bar-charts-in-tableau/ [Accessed: 2024-01-16].

[124] M. Joan Saary. Radar plots: a useful way for presenting multivariate health care data. *Journal of clinical epidemiology*, 61(4):311–317, 2008. ISSN 0895-4356.

[125] Ramik Sadana, Timothy Major, Alistair Dove, and John Stasko. OnSet: A Visualization Technique for Large-scale Binary Set Data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1993–2002, 2014. doi: 10.1109/TVCG.2014.2346249.

[126] Shabana Sangli, Gurminder Kaur, and Bijaya B. Karki. Star plot visualization of ultrahigh dimensional multivariate data. In *Proc. International Conference on Advances in Big Data Analytics*, pages 91–97, 2016.

[127] Shalom H Schwartz, Jan Cieciuch, Michele Vecchione, Eldad Davidov, Ronald Fischer, Constanze Beierlein, Alice Ramos, Markku Verkasalo, Jan-Erik Lönnqvist, Kursad Demirutku, et al. Refining the theory of basic individual values. *Journal of personality and social psychology*, 103(4):663, 2012. doi: 10.1037/a0029393.

[128] SDG Index. Sustainable Development Report 2023: Bangladesh, 2020. URL `https://www.tbsnews.net/economy/sdg-index-2023-bangladesh-ahead-india-pakistan-654006`.

[129] Lin Shao, Timo Schleicher, Michael Behrisch, Tobias Schreck, Ivan Sipiran, and Daniel A. Keim. Guiding the Exploration of Scatter Plot Data Using Motif-Based Interest Measures. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8, 2015. doi: 10.1109/BDVA.2015.7314294.

[130] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proc. 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE, 1996. doi: 10.1109/VL.1996.545307.

[131] Drew Skau and Robert Kosara. Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. *Computer Graphics Forum*, 35(3):121–130, 2016. doi: https://doi.org/10.1111/cgf.12888.

[132] Christian Stab and Iryna Gurevych. Recognizing the Absence of Opposing Arguments in Persuasive Essays. In *Proc. of the Third Workshop on Argument Mining (ArgMining2016)*, pages 113–118, 2016. doi: 10.18653/v1/W16-2813.

[133] Christian Stab and Iryna Gurevych. Argument Annotated Essays (version 2), 2017. URL https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/2422.

[134] Stanley S Stevens. On the psychophysical law. *Psychological review*, 64 (3):153–181, 1957. doi: 10.1037/h0046162.

[135] Daniel Suthers, Arlene Weiner, John Connelly, and Massimo Paolucci. Belvedere: Engaging students in critical discussion of science and public policy issues. In *Proc. of the 7th World Conference on Artificial Intelligence in Education*, pages 266–273, 1995.

[136] Justin Talbot, Vidya Setlur, and Anushka Anand. Four experiments on the perception of bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2152–2160, 2014. doi: 10.1109/TVCG.2014.2346320.

[137] Soon Tee Teoh and Kwan-Liu Ma. PaintingClass: Interactive Construction, Visualization and Exploration of Decision Trees. In *Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 667–672. ACM, 2003. doi: 10.1145/956750.956837.

[138] J. Thompson, Z. Liu, W. Li, and J. Stasko. Understanding the Design Space and Authoring Paradigms for Animated Data Graphics. *Computer Graphics Forum*, 39(3):207–218, 2020. ISSN 0167-7055. doi: 10.1111/cgf.13974.

[139] Sandra A. Thompson and William C. Mann. Rhetorical structure theory. *IPRA Papers in Pragmatics*, 1(1):79–105, 1987. doi: 10.1075/iprapip.1.1.03tho.

[140] Stephen E. Toulmin. *The Uses of Argument*. Cambridge University Press, 2012. ISBN 9780521827485. doi: 10.1017/CBO9780511840005.

[141] Manolis Tzagarakis and Nikos Karacapilidis. On the exploitation of semantic types in the visualization of complex argumentative discourses. In *Proc. of the 2nd International Workshop on Intelligent Exploration of Semantic Data*, page 3, 2013. doi: 10.1145/2462197.2462200.

[142] F. H. van Eemeren, R. Grootendorst, Arnolda Francisca Snoeck Henkemans, and J. Anthony Blair. *Fundamentals of argumentation*

theory: A handbook of historical backgrounds and contemporary developments. Routledge, Taylor & Francis Group, New York, 2009. ISBN 0805818618.

[143] Tim van Gelder. The rationale for Rationale™. *Law, Probability & Risk*, 6(1-4):23–42, 2007. doi: 10.1093/lpr/mgm032.

[144] Frank van Ham and Adam Perer. "Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest. *IEEE Transactions on Visualization and Computer Graphics*, 15 (6):953–960, 2009. doi: 10.1109/TVCG.2009.108.

[145] Christian van Onzenoodt, Pere-Pau Vázquez, and Timo Ropinski. Out of the Plane: Flower versus Star Glyphs to Support High-Dimensional Exploration in Two-Dimensional Embeddings. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5468–5482, 2023. doi: 10.1109/TVCG.2022.3216919.

[146] F. B. Viegas and M. Smith. Newsgroup Crowds and AuthorLines: visualizing the activity of individuals in conversational cyberspaces. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 10 p–, 2004. doi: 10.1109/HICSS.2004.1265288.

[147] Manuela Waldner, Alexandra Diehl, Denis Gračanin, Rainer Splechtna, Claudio Delrieux, and Krešimir Matković. A comparison of radial and linear charts for visualizing daily patterns. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1033–1042, 2019. doi: 10.1109/TVCG.2019.2934784.

[148] Walter Sinnott-Armstrong and Robert J. Fogelin. *Understanding Arguments: An Introduction to Informal Logic*. Cengage Learning, 9 edition, 2013.

[149] D. N. Walton. *Argumentation Schemes for Presumptive Reasoning*. L. Erlbaum Associates, 1996.

[150] Douglas Walton. *Fundamentals of Critical Argumentation*. Cambridge University Press, 2012. ISBN 9780521823197. doi: 10.1017/CBO9780511807039.

[151] Wang Yunhai, Han Fubo, Zhu Lifeng, Deussen Oliver, and Chen Baoquan. Line Graph or Scatter Plot? Automatic Selection of Methods

for Visualizing Trends in Time Series. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1141–1154, 2018. doi: 10.1109/TVCG.2017.2653106.

[152] Weikai Yang, Xiting Wang, Jie Lu, Wenwen Dou, and Shixia Liu. Interactive Steering of Hierarchical Clustering. *IEEE Transactions on Visualization and Computer Graphics*, 27(10):3953–3967, 2021. doi: 10.1109/tvcg.2020.2995100.

[153] Mitchell Whitelaw. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly*, 9(1), 2015.

[154] William S. Cleveland and Robert McGill. The Many Faces of a Scatterplot. *Journal of the American Statistical Association*, 79:807–822, 1984.

[155] Yifan Wu, Joseph M. Hellerstein, and Arvind Satyanarayan. B2: Bridging code and interactive visualization in computational notebooks. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 152–165, 2020.

[156] Yingcai Wu, Zhutian Chen, Guodao Sun, Xiao Xie, Nan Cao, Shixia Liu, and Weiwei Cui. StreamExplorer: A Multi-Stage System for Visually Exploring Events in Social Streams. *IEEE Transactions on Visualization and Computer Graphics*, 24(10):2758–2772, 2018. doi: 10.1109/TVCG.2017.2764459.

[157] Cindy Xiong, Vidya Setlur, Benjamin Bach, Eunyee Koh, Kylie Lin, and Steven Franconeri. Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):955–965, 2021. doi: 10.1109/TVCG.2021.3114823.

[158] Jin Xu, Yubo Tao, Hai Lin, Rongjie Zhu, and Yuyu Yan. Exploring controversy via sentiment divergences of aspects in reviews. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 240–249. IEEE, 2017. doi: 10.1109/PACIFICVIS.2017.8031600.

[159] Tariq Yousef and Stefan Jänicke. A Survey of Text Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1149–1159, 2021. doi: 10.1109/TVCG.2020.3028975.

[160] Xiaoru Yuan, Peihong Guo, He Xiao, Hong Zhou, and Huamin Qu. Scattering Points in Parallel Coordinates. *IEEE Transactions on Visu-*

alization and Computer Graphics, 15(6):1001–1008, 2009. doi: 10.1109/TVCG.2009.179.

[161] Amy X. Zhang, Lea Verou, and David Karger. Wikum: Bridging Discussion Forums and Wikis Using Recursive Summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 2082–2096, New York, NY, USA, 2017. Association for Computing Machinery. doi: 10.1145/2998181.2998235.

[162] Jian Zhao, Fanny Chevalier, Christopher Collins, and Ravin Balakrishnan. Facilitating discourse analysis with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2639–2648, 2012. doi: 10.1109/TVCG.2012.226.

[163] Hong Zhou, Weiwei Cui, Huamin Qu, Yingcai Wu, Xiaoru Yuan, and Wei Zhuo. Splatting the Lines in Parallel Coordinates. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'09, pages 759–766, Chichester, GBR, 2009. The Eurographs Association & John Wiley & Sons, Ltd. doi: 10.1111/j.1467-8659.2009.01476.x.

[164] Liang Zhou and Daniel Weiskopf. Indexed-points parallel coordinates visualization of multivariate correlations. *IEEE Transactions on Visualization and Computer Graphics*, 24(6):1997–2010, 2018. doi: 10.1109/TVCG.2017.2698041.