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

Dora Kiesel, Patrick Riehmann, Ines Engelmann, Hanna Ramezani and Bernd Froehlich

***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 communication 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.**

***Index Terms*—Information Visualization, Visual Text Analytics, Text Visualization, Graph Visualization, Multi-Attribute Aggregation, Tree Aggregation, Set-based Visualization**

I. 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 consist sometimes 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 [1]. 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 [2], [3].

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 compare 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. [4], which is limited to binary node attributes and a single a priori aggregation criterion.

Thus, we introduce the Comment Analyzer (Figure 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 communication scientists to draw conclusions about common shapes of comment trees and sequences of deliberative indicators associated with the temporal development of a discussions 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

Dora Kiesel is with the Virtual Reality and Visualization Group, Bauhaus-Universität Weimar, Germany. Email: dora.kiesel@uni-weimar.de

Patrick Riehmann is with the Computer Science and Informatics Department, School of Engineering, Jönköping University, Sweden.

Ines Engelmann is with the Empirical Methods of Communication Science, Friedrich Schiller University Jena, Germany.

Hanna Ramezani is with the German Federal Agency for Civic Education.

Bernd Froehlich is with the Virtual Reality and Visualization Group, Bauhaus-Universität Weimar, Germany.

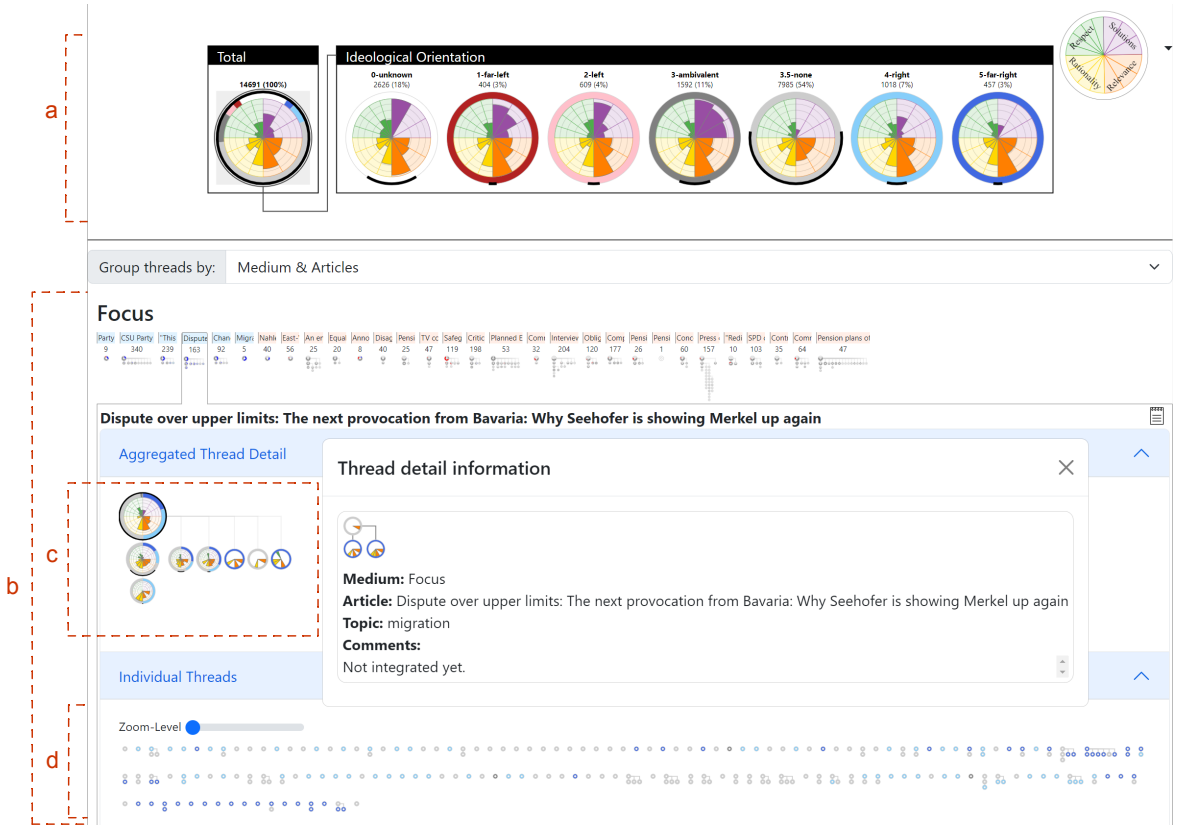


Fig. 1: The Comment Analyzer during an analysis process: In the visual set-shaping component (a), the entire dataset has been split into subsets by ideological orientation, revealing the differences in the individual deliberative indicators. Each individual Comment Glyph shows the ideological orientation 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.

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 II) 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 [5].

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.

II. Related Work

Tree aggregation, visualization and text. Regular tree and graph visualizations are used to convey relationships in communication (Internet Newsgroups [6], (Multi)ConVis [7], [8], Wikum [9]) or words (WordTree [10], PhraseNets [11], Netspeak Wordgraph [12], [13]). See also Yousef and Jänicke's [14]

survey on text alignment, the Dagstuhl report on text visualization [15] or the Text Visualization Browser [16]. Comparing a huge number of trees is not easy. The *DAViewer* [17] compares discourse trees with interactive dendrograms, while Lahmar and Herschel [18] aggregate provenance traces. Liu et al. [19], Bremm et al. [20] and Munzner et al. [21] contrast phylogenetic trees as small multiple or through aggregation. Graham and Kennedy [22] surveyed tree visualizations, stating that aggregation – as our Union Tree – is the most space efficient. Kiesel et al. [4] is closely related, aggregating argument structures of essays, but limited to binary attributes and a fixed aggregation criterion. PansyTree [23] 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 [24], [25], [26], our work relies on round glyphs due to their aesthetics and perceived closure (see Borgo et al. [27] for general rules). Designs for multivariate data are the star glyph [28], [29] and flower glyph [30], [31], [32], 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 [33], (Multi)ConVis [8], [7] and DoSVis [34] dedicate separate views for stance visualization, the StreamExplorer [26] uses a pie layout. Similar to Xu et al. [35], our Comment Glyph employs a donut chart to depict stance. Skau and Kosara [36] found that donut charts and pie charts work similar. Kucher et al. [37] 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. [38] chose Sankey diagrams, we adopt a node-link layout, like Castro and Bertini [39] and PaintingClass [40], 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 [41] 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. [42], [43] and ConVisIT [44]. Topical development and deliberative quality over time can be explored with NEREx [45] or Lexical Episode Plots [46], patterns in a discussion can be found in Jentner et al. [47] or the topic-space of a discussion explored with ConToVi [48]. The iForum [49] 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 [50], [51]. 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. [52] and van Ham and Perer [53] used a similar approach for large graph exploration. Luciani et al. [54] showed the volume details first. And even though our final system adheres to Shneiderman's top-down mantra [55], 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. [4] with a more sophisticated node representation and extended set shaping facilities. The flower glyph [30] 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.

III. 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 I). 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.

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 [2], [3], [56]. 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. [57]

- **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.

TABLE I: 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

IV. 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 V) in two steps. First, we conducted in-depth reviews with the involved experts (see Section VI) to validate existing findings but more importantly investigate new ones (Table II). Later on, as separate stage but with a similar protocol, we did reviews with five independent experts (also communication sciences, described in Section VII).

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 [2], [3], [56]. 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 [55] – 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.

V. Comment Analyzer: Design and Application

The Comment Analyzer (Figure 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" [58], we designed tangible and visually aggregable representatives for the main data items: The Comment Glyph aggregates comment sets, the Union Tree thread sets.

A. 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.

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

Fig. 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.

The glyph (see iterations in Figure 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 I 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 between bars due to its more effective length encoding [59], 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 [60] 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.

B. Aggregating Threads as Union Trees

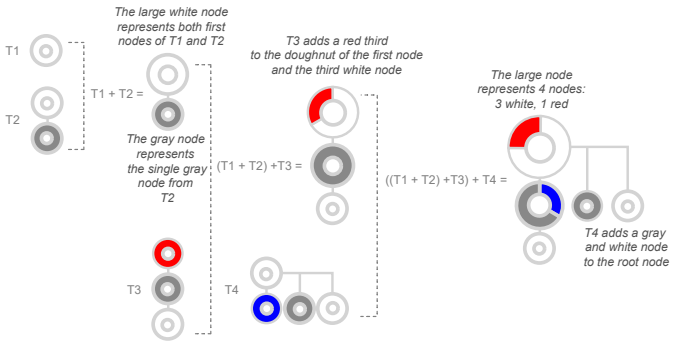


Fig. 3: Construction of a Union Tree: The mechanism superimposes nodes while maintaining structural positions of every tree. The nodes' attributes (here shown as colored donut pieces) are aggregated. New nodes (along with their adjacent edges) were added to the Union Tree. The colors conform with the 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.

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 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 shows how a Union Tree is constructed. Individual threads (T1-4 in Figure 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). 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 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.

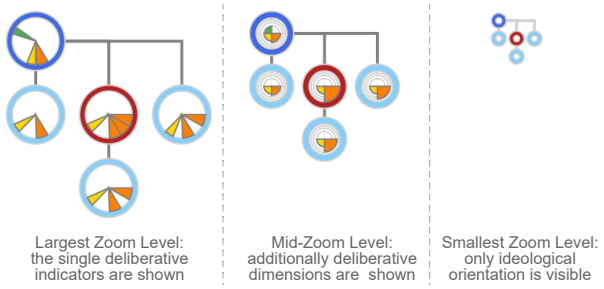


Fig. 4: Single-Comment Glyphs at different semantic zoom levels. They are used to depict the individual threads in Figure 8 and T1-4 in Figure 3.

C. 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 1b and also Figure 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 4 or individual threads in Figure 7).

Opening a Union Tree on demand reveals its highest level of detail, using the full Comment Glyph (see Aggregated Thread Detail in Figure 7) for expressing multiple comments. A Glyph for an individual comment (Figure 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 [5] at 9:08).

D. Interactive Incremental Visual Set-Shaping

The set-shaping component (Figure 1a and 5) enables users to gain insight into the features of different subsets of com-

ments. 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 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.

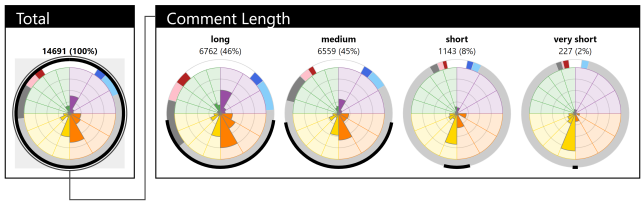
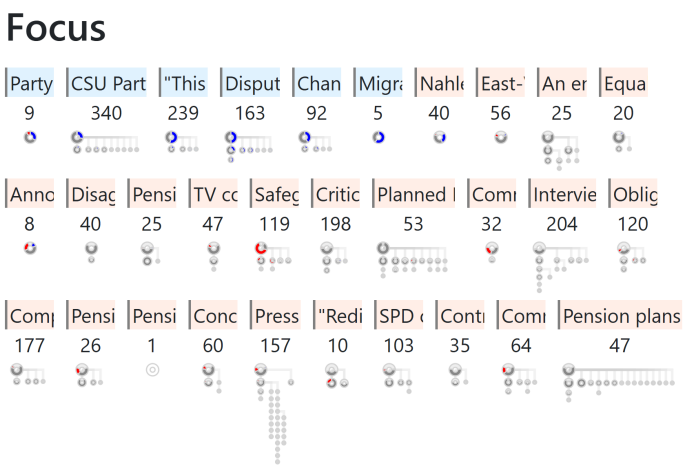


Fig. 5: The visual set-shaping component splits the whole corpus (left) by comment length (long: > 60 words, medium: ≤ 60 words, short: ≤ 15 words, very short: ≤ 5 words). The set's sizes are written above and also encoded as length of the black outer border of each glyph. The split reveals that the 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.

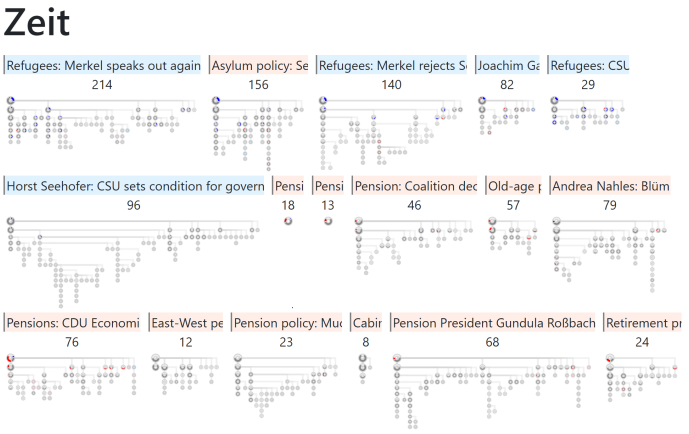
E. Aggregation-based Structure Analysis

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 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 different 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



(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.

Fig. 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).

once. Different **Groupings** can be selected (topics, medium, depth of discussion, etc.) to **Aggregate** the threads. Figure 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 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

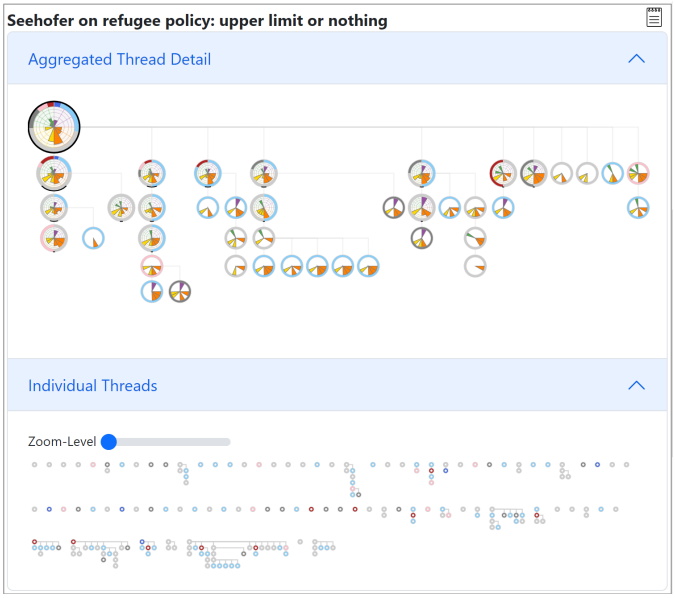


Fig. 7: The Union Tree (Aggregated Thread Detail) in the upper part summarizes the structures and node attributes of all individual comment threads in the lower part – here all comment threads from the German news article "Seehofer about migration politics: Upper limit or nothing" – into an interactive 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 4 for zoom levels.

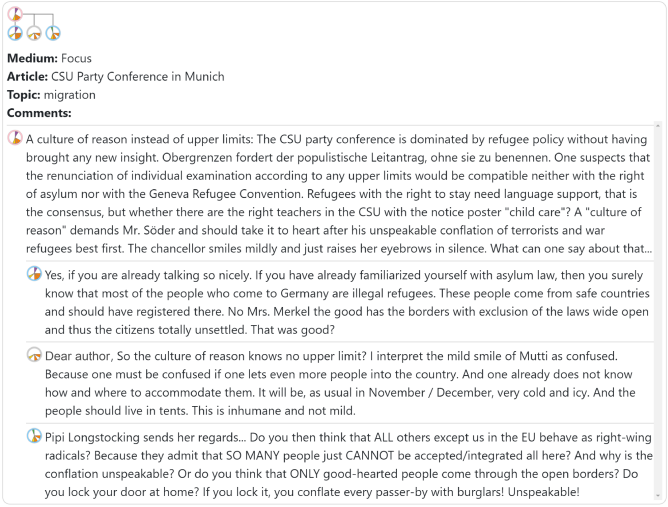


Fig. 8: The thread detail shows full detailed Comment Glyphs, meta information and the actual texts as indented tree. The example shows a discussion about a convention of the German party CSU where upper limits of immigration were discussed. It starts with a justified and topic relevant (2 orange slices) 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.

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 coordination additionally helps when the aggregated thread detail and the individual threads are opened (Figure 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 10) comment as discussed in the expert review section. Details about an individual thread can be opened on demand in a separate floating view (Figure 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.

F. *Interactive Scrapbook*

The scrapbook component (Figure 10) 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 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.

VI. Domain Findings with Associated Experts

For the final reviews we followed the *pair analytics study* approach [61] 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 II). 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 II) 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 III. Table II 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.

A. Domain-Related Findings

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

P3 + P4: Validation of previous findings. In their previous studies, the experts compared single deliberative indicators and quality dimensions 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 11a), 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 11b), 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 9. 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 10 shows such an example 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 10 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.

VII. 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.

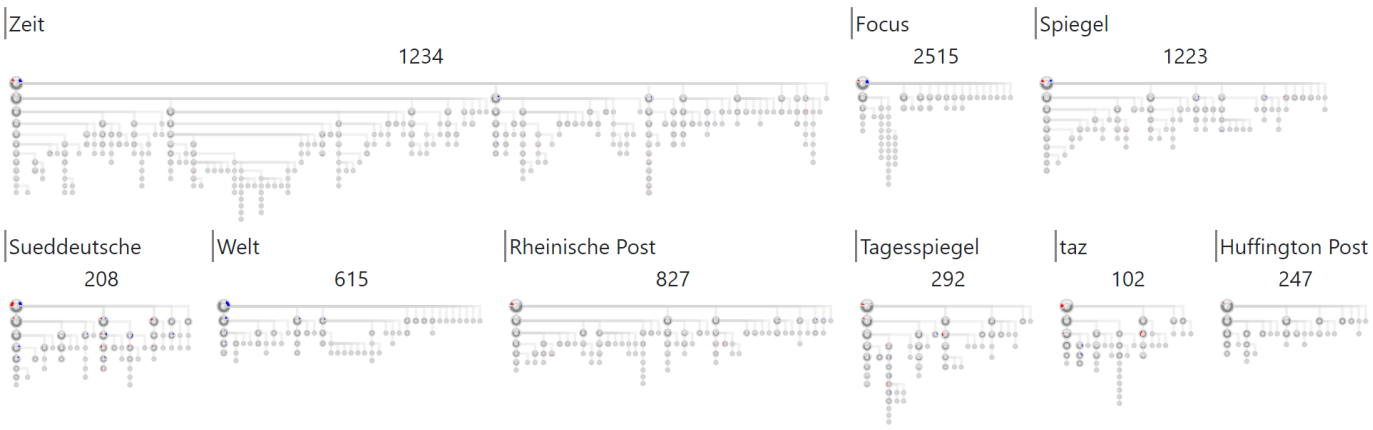


Fig. 9: The Union Trees aggregated up to news platform clearly show the variability of the width and depth of the trees as well as the size of node areas (indicating the number of nodes existing at this position). These observations contributed in part to preliminary reasoning of our experts about the influence of well-designed discussion features provided in the UIs of the news platforms, but will also be subject to future investigation.

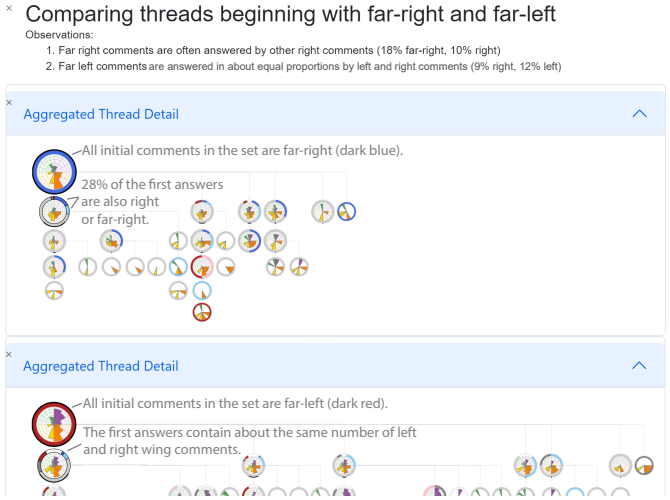


Fig. 10: Example scrapbook during an analysis. The scientist compares the proportions of ideological orientation between threads starting far-right (dark blue) and far-left (dark red) and notes down their observations.

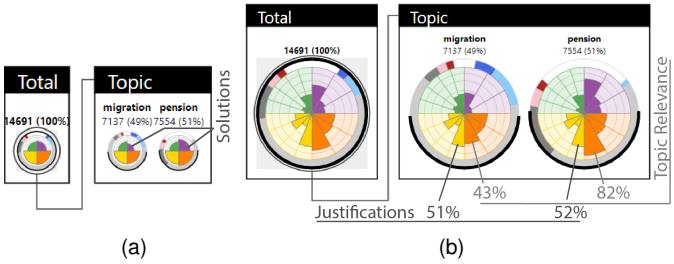


Fig. 11: Findings about quality dimensions and deliberative indicators over two topics: **(a)** The medium zoom level shows how many comments have indicators in each of the four quality dimensions. The size difference of purple quadrants reveals more proposals of solutions to comments about pensions 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.

Again, we followed the *pair analytics study* [61] 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 II 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.

A. 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* →

TABLE II: 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]	✗

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 com-

parison) 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.

B. 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 about 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 about 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 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.

VIII. 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.

IX. 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.

ACKNOWLEDGMENTS

This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 16DHBKI083 (project SKILL).

REFERENCES

[1] P. Aragón, V. Gómez, and A. Kaltenbrunner, "Detecting Platform Effects in Online Discussions," *Policy & Internet*, vol. 9, no. 4, pp. 420–443, 2017.

[2] H. Marzinkowski and I. Engelmann, "Rational-critical user discussions: how argument strength and the conditions set by news organizations are linked to (reasoned) disagreement," *Digital Journalism*, vol. 10, no. 3, pp. 433–451, 2022.

[3] I. Engelmann, H. Marzinkowski, and K. Langmann, "Salient deliberative norm types in comment sections on news sites," *New Media & Society*, vol. 26, no. 2, pp. 921–940, 2024.

[4] D. Kiesel, P. Riehmann, H. Wachsmuth, B. Stein, and B. Froehlich, "Visual Analysis of Argumentation in Essays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1139–1148, 2021.

[5] D. Kiesel, "Video on the main concepts and workflows of the Comment Analyser," 2024. [Online]. Available: https://drive.google.com/file/d/1qtb_KfjI4tiRrMWjUy4ul7GI2-S5WiTo/view?usp=drive_link

[6] F. B. Viégas and M. Smith, "Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals in Conversational Cyberspaces," in *Proc. of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)*. IEEE Computer Society, 2004, p. 40109.2.

[7] E. Hoque and G. Carenini, "ConVis: A Visual Text Analytic System for Exploring Blog Conversations," *Computer Graphics Forum*, vol. 33, no. 3, pp. 221–230, 2014.

[8] ——, "MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations," in *Proc. of the 21st International Conference on Intelligent User Interfaces*. ACM, 2016, pp. 96–107.

[9] A. X. Zhang, L. Verou, and D. Karger, "Wikum: Bridging Discussion Forums and Wikis Using Recursive Summarization," in *Proc. of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2017, p. 2082–2096.

[10] M. Wattenberg and F. B. Viégas, "The Word Tree, an Interactive Visual Concordance," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1221–1228, 2008.

[11] F. van Ham, M. Wattenberg, and F. B.Viégas, "Mapping Text with Phrase Nets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 06, pp. 1169–1176, 2009.

[12] P. Riehmann, H. Gruendl, B. Froehlich, M. Potthast, M. Trenkmann, and B. Stein, "The NETSPEAK WORDGRAPH: Visualizing keywords in context," in *2011 IEEE Pacific Visualization Symposium*, 2011, pp. 123–130.

[13] P. Riehmann, H. Gruendl, M. Potthast, M. Trenkmann, B. Stein, and B. Froehlich, "WORDGRAPH: Keyword-in-Context Visualization for NETSPEAK's Wildcard Search," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1411–1423, 2012.

[14] T. Yousef and S. Jänicke, "A Survey of Text Alignment Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1149–1159, 2021.

[15] C. Collins, A. Fokkens, A. Kerren, C. Weaver, and A. Chatzimparmpas, "Visual Text Analytics (Dagstuhl Seminar 22191)," *Dagstuhl Reports*, vol. 12, no. 5, pp. 37–91, 2022. [Online]. Available: https://drops.dagstuhl.de/opus/volltexte/2022/17443

[16] K. Kucher and A. Kerren, "Text visualization techniques: Taxonomy, visual survey, and community insights," in *2015 IEEE Pacific Visualization Symposium (PacificVis)*, 2015, pp. 117–121.

[17] J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan, "Facilitating discourse analysis with interactive visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2639–2648, 2012.

[18] H. B. Lahmar and M. Herschel, "Structural summaries for visual provenance analysis," in *11th International Workshop on Theory and Practice of Provenance (TaPP 2019)*. USENIX Association, Jun. 2019, p. 2.

[19] Z. Liu, S. H. Zhan, and T. Munzner, "Aggregated Dendrograms for Visual Comparison between Many Phylogenetic Trees," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 9, pp. 2732–2747, 2020.

[20] S. Bremm, T. von Landesberger, M. Heß, T. Schreck, P. Weil, and K. Hamacherk, "Interactive visual comparison of multiple trees," in *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2011, pp. 31–40.

[21] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou, "TreeJuxtaposer: scalable tree comparison using Focus+Context with guaranteed visibility," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, pp. 453–462, 2003.

[22] M. Graham and J. Kennedy, "A survey of multiple tree visualisation," *Information Visualization*, vol. 9, no. 4, pp. 235–252, 2010.

[23] Y. Dong, A. Fauth, M. Huang, Y. Chen, and J. Liang, "PansyTree: Merging Multiple Hierarchies," in *2020 IEEE Pacific Visualization Symposium (PacificVis)*, 2020, pp. 131–135.

[24] S. Fu, Y. Wang, Y. Yang, Q. Bi, F. Guo, and H. Qu, "VisForum: A Visual Analysis System for Exploring User Groups in Online Forums," *ACM Transactions on Interactive Intelligent Systems*, vol. 8, no. 1, pp. 1–21, 2018.

[25] M. Glueck, M. P. Naeini, F. Doshi-Velez, F. Chevalier, A. Khan, D. Wigdor, and M. Brudno, "PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 371–381, 2018.

[26] Y. Wu, Z. Chen, G. Sun, X. Xie, N. Cao, S. Liu, and W. Cui, "StreamExplorer: A Multi-Stage System for Visually Exploring Events in Social Streams," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 10, pp. 2758–2772, 2018.

[27] R. Borgo, J. Kehrer, D. H. S. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications," in *Eurographics 2013 - State of the Art Reports*. The Eurographics Association, 2013, pp. 39 – 63.

[28] J. Fuchs, P. Isenberg, A. Bezerianos, F. Fischer, and E. Bertini, "The influence of contour on similarity perception of star glyphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2251–2260, 2014.

[29] T. Opach, S. Popelka, J. Dolezalova, and J. K. Rød, "Star and polyline glyphs in a grid plot and on a map display: which perform better?" *Cartography and Geographic Information Science*, vol. 45, no. 5, pp. 400–419, 2018.

[30] M. Keck, D. Kammer, T. Gründer, T. Thom, M. Kleinsteuber, A. Maasch, and R. Groh, "Towards Glyph-Based Visualizations for Big Data Clustering," in *Proc. of the 10th International Symposium on Visual Information Communication and Interaction*. ACM, 2017, p. 129–136.

[31] C. van Onzenoodt, P.-P. Vázquez, and T. 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*, vol. 29, no. 12, pp. 5468–5482, 2023.

[32] I. Cho, R. Wesslen, S. Volkova, W. Ribarsky, and W. 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)*, 2017, pp. 25–35.

[33] R. M. Martins, V. Simaki, K. Kucher, C. Paradis, and A. Kerren, "StanceXplore: Visualization for the Interactive Exploration of Stance in Social Media," in *Proc. of 2nd Workshop on Visualization for the Digital Humanities*, 2017, pp. 1–5.

[34] K. Kucher, C. Paradis, and A. 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*. SciTePress, 2018, pp. 168–175.

[35] J. Xu, Y. Tao, H. Lin, R. Zhu, and Y. Yan, "Exploring Controversy via Sentiment Divergences of Aspects in Reviews," in *2017 IEEE Pacific Visualization Symposium (PacificVis)*, 2017, pp. 240–249.

[36] D. Skau and R. Kosara, "Arcs, angles, or areas: Individual data encodings in pie and donut charts," *Computer Graphics Forum*, vol. 35, no. 3, pp. 121–130, 2016.

[37] K. Kucher, C. Paradis, and A. Kerren, "The State of the Art in Sentiment Visualization," *Computer Graphics Forum*, vol. 37, no. 1, pp. 71–96, 2018.

[38] W. Yang, X. Wang, J. Lu, W. Dou, and S. Liu, "Interactive Steering of Hierarchical Clustering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 10, pp. 3953–3967, 2021.

[39] F. D. Castro and E. 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)*. ACM, 2019, pp. 1 – 5.

[40] S. T. Teoh and K.-L. 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. ACM, 2003, p. 667–672.

[41] R. Sadana, T. Major, A. Dove, and J. Stasko, "OnSet: A Visualization Technique for Large-Scale Binary Set Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1993–2002, 2014.

[42] M. El-Assady, A. Hautli-Janisz, V. Gold, M. Butt, K. Holzinger, and D. Keim, "Interactive Visual Analysis of Transcribed Multi-Party Discourse," in *Proc. of ACL 2017, System Demonstrations*. Association for Computational Linguistics, 2017, pp. 49–54.

[43] M. El-Assady, R. Sevastjanova, D. Keim, and C. Collins, "ThreadReconstructor: Modeling Reply-Chains to Untangle Conversational Text through Visual Analytics," *Computer Graphics Forum*, vol. 37, no. 3, pp. 351–365, 2018.

[44] E. Hoque and G. Carenini, "ConVisIT: Interactive Topic Modeling for Exploring Asynchronous Online Conversations," in *Proc. of the 20th International Conference on Intelligent User Interfaces*. ACM, 2015, p. 169–180.

[45] M. El-Assady, R. Sevastjanova, B. Gipp, D. Keim, and C. Collins, "NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations," *Computer Graphics Forum*, vol. 36, no. 3, pp. 213–225, 2017.

[46] V. Gold, C. Rohrdantz, and M. El-Assady, "Exploratory Text Analysis using Lexical Episode Plots," in *Eurographics Conference on Visualization (EuroVis)*, 2015, pp. 85–89.

[47] W. Jentner, M. El-Assady, B. Gipp, and D. A. Keim, "Feature Alignment for the Analysis of Verbatim Text Transcripts," in *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2017, pp. 13 – 17.

[48] M. El-Assady, V. Gold, C. Acevedo, C. Collins, and D. Keim, "ConToVi: Multi-Party Conversation Exploration using Topic-Space Views," *Computer Graphics Forum*, vol. 35, no. 3, pp. 431–440, 2016.

[49] S. Fu, J. Zhao, W. Cui, and H. Qu, "Visual Analysis of MOOC Forums with iForum," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 201–210, 2017.

[50] Y. Wu, J. M. Hellerstein, and A. Satyanarayan, "B2: Bridging code and interactive visualization in computational notebooks," in *Proc. of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2020, pp. 152–165.

[51] R. Kosara, "Notebooks for data analysis and visualization: Moving beyond the data," *IEEE Computer Graphics and Applications*, vol. 43, no. 1, pp. 91–96, 2023.

[52] B. Lee, C. Parr, C. Plaisant, B. Bederson, V. Veksler, W. Gray, and C. Kotfila, "TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1414–1426, 2006.

[53] F. van Ham and A. Perer, "Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 953–960, 2009.

[54] T. Luciani, A. Burks, C. Sugiyama, J. Komperda, and G. E. Marai, "Details-First, Show Context, Overview Last: Supporting Exploration of Viscous Fingers in Large-Scale Ensemble Simulations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, p. 1225–1235, 2019.

[55] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *Proc. 1996 IEEE Symposium on Visual Languages*. IEEE, 1996, pp. 336–343.

[56] H. Marzinkowski and I. Engelmann, "Die Wirkung 'guter' Argumente," *Publizistik*, vol. 63, no. 2, pp. 269–287, 2018.

[57] M. Dörk, S. Carpendale, and C. Williamson, "The Information Flaneur: A Fresh Look at Information Seeking," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, p. 1215–1224.

[58] M. Whitelaw, "Generous Interfaces for Digital Cultural Collections," *Digital Humanities Quarterly*, vol. 9, no. 1, 2015.

[59] W. S. Cleveland and R. McGill, "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531–554, 1984.

[60] S. S. Stevens, "On the psychophysical law," *Psychological review*, vol. 64, no. 3, pp. 153–181, 1957.

[61] L. T. Kaastra and B. Fisher, "Field Experiment Methodology for Pair Analytics," in *Proc. of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. ACM, 2014, p. 152–159.

BIOGRAPHY SECTION

Dora Kiesel is a doctoral candidate at the Virtual Reality and Visualization Research Group at Bauhaus-Universität Weimar. Her research interests include visual text analytics, digital humanities and embedding technologies.

Patrick Riehmann is with Jönköping University as Assistant Professor for Information and Data Visualization. His research covers the topics Visual Analytics and HCI in particular for analyzing multivariate time series and investigating text and argumentation structures. He is also interested in visual decision and configuration support as well as adapting Information Visualization paradigms for very large multi-touch displays. More information at Visual Metaphors (http://riehmann.net/).

Ines Engelmann is a full professor at the Institute of Communication Science at Friedrich Schiller University Jena, Germany (https://www.fsv.uni-jena.de/en/15345/empirical-methods-of-communication-science). Her research interests focus on journalistic news construction (e.g., news values, political value frames, characteristics of politicians), audience participation in user comment sections (e.g., discourse architectures, political motives, discussion quality) and alternative online political media (e.g., news routines, populism).

Hanna Ramezani is an advisor at the German Federal Agency for Civic Education (Bundeszentrale für politische Bildung) in the department for civic education and social media. Her scientific interests include online discourses, online deliberation, and political communication.

Bernd Froehlich is a full professor with the Computer Science Department at Bauhaus-Universität Weimar and head of the Virtual Reality and Visualization Research Group (www.uni-weimar.de/medien/vr). His research interests include real-time rendering, visualization, 3D user interfaces, social virtual reality and immersive telepresence.