Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

netflower: Dynamic Network Visualization for Data Journalists

C. Stoiber[1], A. Rind[1], F. Grassinger[1], R. Gutounig[2], E. Goldgruber[2], M. Sedlmair[3], Š. Emrich[4] and W. Aigner[1]

[1]St. Poelten University of Applied Sciences, Austria
[2]FH JOANNEUM – University of Applied Sciences, Graz, Austria
[3]University of Stuttgart, Germany
[4]drahtwarenhandlung, Landsiedl Popper OG, Austria

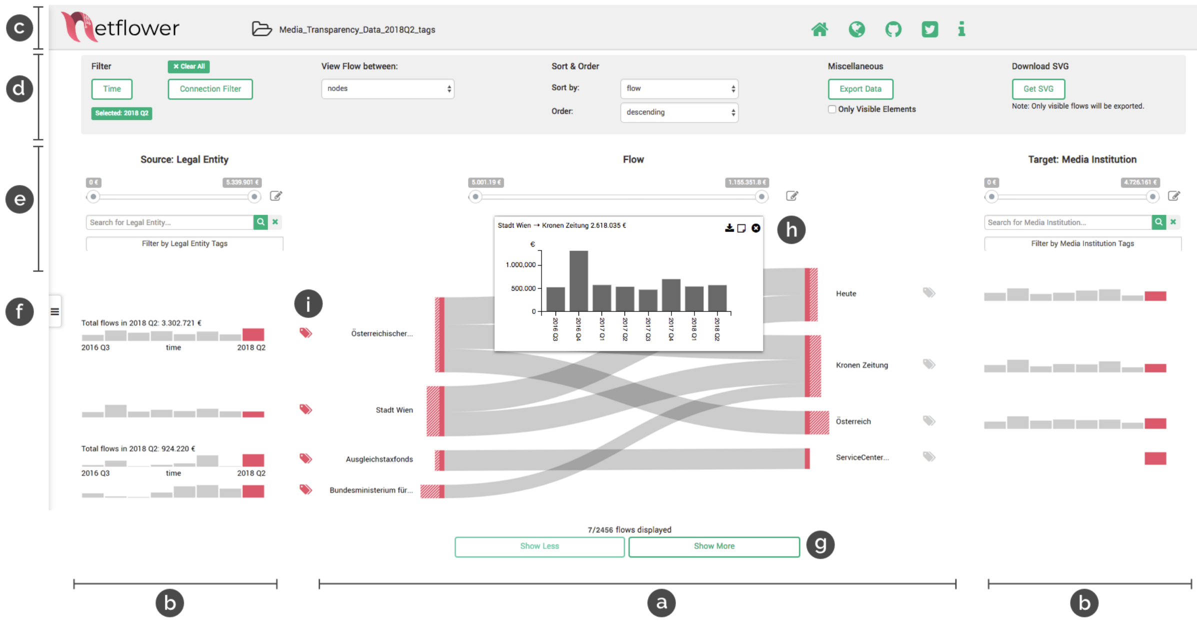

Figure 1: *The visual exploration tool netflower supports journalists in investigating quantitative flows in dynamic network data for story-finding. The example above shows the media transparency dataset which tracks quarterly money flows between government institutions and media outlets. The main visualization is a Sankey diagram (a) showing the flows between legal entities and media institutions. Bar chart sparklines on both sides (b) give a temporal overview of outgoing (source nodes) and incoming money (target nodes). To load more data than visible at first glance, the "show more" and "show less" buttons (g) can be used. An on-demand bar chart (h) is provided to show the temporal evolution of one particular flow. Tags can be assigned to the nodes on both sides (i) and then filtered for. Below the header area (c), netflower provides functions for filtering, sorting and ordering (d, e). Moreover, a notebook is integrated into a sidebar (f) for investigative provenance.*

Abstract

Journalists need visual interfaces that cater to the exploratory nature of their investigative activities. In this paper, we report on a four-year design study with data journalists. The main result is netflower, a visual exploration tool that supports journalists in investigating quantitative flows in dynamic network data for story-finding. The visual metaphor is based on Sankey diagrams and has been extended to make it capable of processing large amounts of input data as well as network change over time. We followed a structured, iterative design process including requirement analysis and multiple design and prototyping iterations in close cooperation with journalists. To validate our concept and prototype, a workshop series and two diary studies were conducted with journalists. Our findings indicate that the prototype can be picked up quickly by journalists and valuable insights can be achieved in a few hours. The prototype can be accessed at: http://netflower.fhstp.ac.at/

CCS Concepts

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

1. Introduction and background

We live in a world in which it is increasingly important to understand complex socio-economic and ecological phenomena to facilitate well-informed decisions. Journalists play an important role in this endeavor by uncovering hidden patterns and relationships to inform, enlighten and entertain. With the ever-growing amount and availability of data, it becomes crucial for journalists to use elements of data science in their work. This trend has led to the advent of the emerging field of Data-Driven Journalism (DDJ) which involves computer-supported data-based reasoning as well as interactive visualization [AGO*17]. Although DDJ has received attention from different communities and has found its way into a number of well-known news organizations such as the New York Times and the Guardian, the majority of journalists still face significant obstacles inhibiting the utilization of data for their work. Four years ago, we set out to work with this user group with the goal of understanding their needs and interests better, and to investigate the potential value that interactive visualization can contribute to their work. In this paper, we report on a design study from this process.

Flows in dynamic networks: Our design study focuses on analyzing dynamic network data which we found to be a very common data type faced by our target users. Journalists may, for example, wish to analyze a social network in terms of how communication (edges in the network) and important actors (nodes) change over time. Often, these networks additionally include quantitative flows between nodes, such as money flows between organizations, migration flows between countries, or transfer payments between sports clubs.

These flows can be conceptualized as time-dependent edge weights in a directed network [RPNA16]. Edge weights are typically non-negative quantitative attributes that count or sum up features from individual events occurring in the course of a year, quarter, or month. The network is dynamic both in terms of its structure (nodes and edges can appear or disappear over time) and its quantitative flows (weights changing over time) [vLKS*11]. The time-oriented aspect of the data can be characterized as instants on a discrete, interval-based, linear time domain with granularities for quarter and year [AMST11]. Some of these networks are bipartite, i.e., their nodes can be divided into two disjoint sets and each edge connects nodes of different sets. The Austrian media transparency dataset (MTD) distinguishes government nodes that spend money from media nodes that receive money. In contrast, a migration dataset can have the same country as source and target of migration flows. For some datasets, it is possible to further break up flows into different categories by attribute, such as the type of spending or the legal status of a migrant.

In the following, we will use the Austrian media transparency dataset (MTD) [RTR18] as a representative real-world example to illustrate the design of *netflower*. MTD means open government data that reports money spent by Austrian governmental bodies on advertisement and media sponsorship. Currently, data are available for 8 quarters (Q3/2016–Q2/2018) comprising 18,302 quarterly money flows (in total, without missing values). These flows connect 825 different legal entities (e.g., federal ministries, cities, economic chambers, government-owned companies) to 1,545 distinct media institutions (e.g., newspaper, TV, radio, online). For a deeper exploratory data analysis of the MTD, see [RPNA16]. We selected the MTD as a representative example because (i) it is publicly available, (ii) it has a non-trivial scale, (iii) journalists are familiar with the topic, and (iv) it is regularly updated which makes it possible to add current stories during and after the design study project possible.

Design study contribution: Although such dynamic, weighted and directed graph structures can be found in many data-journalistic areas (and also in many other application domains), we found that the current tool support for these is scarce. Surprisingly, there are not many visualization systems that do, in fact, support the visual exploration of dynamic networks at a more general level (see Sec. 7). General tools such as Tableau [tab] mainly focus on multivariate data, while standard network visualization tools such as Gephi [gep] do not support dynamic networks well enough.

To help bridge this gap, we report on the design, development, and evaluation of *netflower* (dynamic network flow explorer) to support data journalists in their analysis tasks. Over the last four years, we collaborated closely with different groups of target users, iteratively forming a better understanding of the problem as well as designing and evaluating the *netflower* tool to address it. Methodologically, we specifically relied on interviews and design workshops, which we thought would work well for our time-pressured target audience. Through this process, we eventually arrived at a design that leverages (i) Sankey diagrams [Sch08], as an easy-to-understand idiom for weighted dynamic graphs, (ii) a set of well-crafted analysis workflows tailored towards journalists' needs, as well as (iii) first onboarding techniques that help users get started quickly.

In summary, the main contributions of this work are:

- a problem characterization with data and task abstractions (see Sec. 3)
- the design of *netflower* which is described in Sec. 4; a visual exploration tool that enables journalists to understand phenomena that can be phrased as dynamic, weighted graphs,
- the results of four workshops and two diary studies in which different versions of the tool were used and tested with journalists (see Sec. 6), and
- a set of general lessons learned which we derived from our project, such as supporting journalists in visual exploration and story-finding, and using empowerment workshops as an appropriate method to evaluate visualization designs with journalists (see Sec. 8).

2. Methods

As is common for design studies, we followed a process of iterative design and evaluation to address a particular domain problem by involving collaborators and users from the domain [SMM12]. We decided to validate the *netflower* prototype based on *workshops* and *diary studies*. The workshops were chosen as an incentive for journalists to reach out to busy domain experts. Diary studies were aimed at evaluating the netflower system in a more realistic setting, while again considering the experts' tight schedule. A timeline illustrating *netflower's* development and the iterative validation

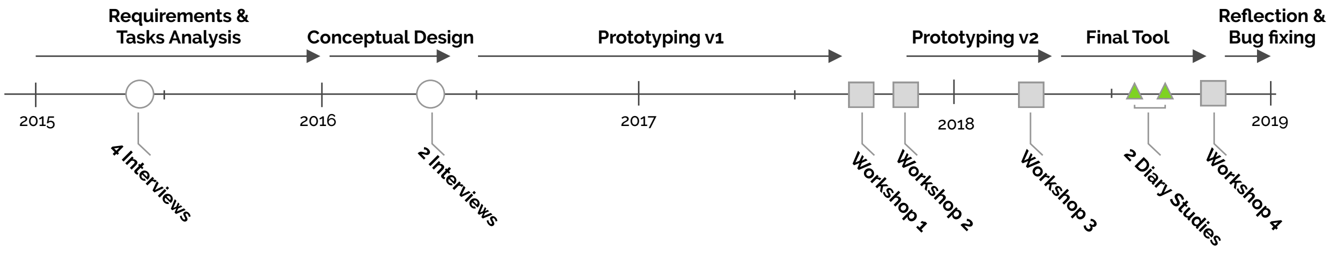

Figure 2: *Timeline of netflower's development and validation phases from 2015 to end of 2018*

phases is shown in Figure 2. After beginning with the requirement and task analysis in 2015 with four semi-structured interviews in May, we produced paper prototypes and mockups. We revised these based on a follow-up interview with two of the domain experts. In 2016, we started to develop a functional prototype *v1* and tested this prototype in two workshops involving two students and three journalists. After the first evaluation phase, we integrated the feedback we received and further improved the prototype, resulting in prototype *v2*. A third workshop (six participants) was then conducted to test *v2*. Between the third and fourth workshop with journalists, we conducted two diary studies, using *v2* simultaneously. Finally, the finished tool was evaluated in a workshop setting (three journalists). The participants of all workshops differed from each other, from the interviewees, and the diary study participants.

Additionally, we used the SUS questionnaire [Bro13] as a standardized usability score for our evolving prototype but found that the results were not conclusive due to the small number of participants for this quantitative instrument and the ensuing limited generalizability. Hence, we concentrated on qualitative feedback of our domain experts.

In the following section, we report on the results of this iterative design process. We present the problem characterization and abstraction we derived, the design decisions that governed our process, and the evaluation results from the studies.

3. Problem characterization and abstraction

In the first phase of the *netflower* design study, we conducted interviews [LFH09] with four data journalists to obtain a good understanding of our target group's analysis and workflow requirements.

3.1. Method and participants

To collect a first set of design requirements, we conducted semi-structured interviews [LFH09] with four data journalists (I1–4) in 2015. The interviews were approximately one hour in duration and were performed either in person or via video call. All of the surveyed journalists were male and had more than 10 years of experience in journalism. In addition to their journalistic activities, three of them teach journalism or data journalism at different universities in Austria. Throughout the interviews, we used the MTD to guide participants through a potential scenario. In this context, we investigated how the journalistic process is structured, how it is documented, which tools are used, and what potential problems accompany these tools. After that, we went into more detail when it comes to the example dataset and explored which aspects are of special interest (e.g., temporal and relational). Finally, we specifically asked questions about visualization approaches, i.e., which techniques and tools the journalists are familiar with based on a set of examples we prepared.

3.2. Results

All four interviewees knew the MTD but only one of them had actively worked with the dataset towards creating a news article. A preliminary analysis has been reported in [NRA*16].

According to our interviewees, data journalists in Austria have basic computer literacy as well as basic statistical knowledge (cp. [Bra11]). The exploration and research phase was organized depending on the journalists' personal preferences and skills. The interviews indicated that journalists are comfortable with combining different tools such as Microsoft Excel, Google Spreadsheets for data analysis and other tools such as Evernote [eve] for storing notes, which were found during the exploration or research phase. Furthermore, the research documentation has been reported to be in both electronic and paper forms. A commonly followed standard workflow or specific user description could not be identified on the basis of the conducted interviews. Due to the investigative character of their work, the interviewed journalists have to deal with a wide range of data sources originating from various other domains. However, deadlines and lack of time (and resources) for extensive analysis are also limitations to data-intensive news work. When it comes to visualization techniques, standard business charts such as bar charts, line plots or pie charts were mentioned as being known and used. The journalists stated that they prefer easy-to-use interfaces and visualizations. The journalists reported that they usually carry out the research phase on a news story alone. The next steps of designing infographics and drafting the final article are then mostly done in teams of two to three people.

The interviewed journalists either start from a story idea and subsequently search data to support the evidence, or they start from data and develop a story idea revolving around their findings. Before writing a story, however, they would conduct interviews because it is their journalistic responsibility to compare information from multiple sources. In this process, data analysis helps them to identify possible interview partners. One interviewee also mentioned that data-based insights help him to get more time for investigation approved.

In general, data journalists browse their data to discover outliers and other anomalies. Exploring the MTD, a possible starting point would be a legal entity or media institution that shows abrupt changes in the sum of money spent or received, or a varying number of edges, e.g., *"why was so much money spent?"* (I3). For some stories, they would apply filters to focus on the geographic region where their newspaper is based. Or they would focus on a time period: *"Look at the numbers, when was something introduced and when was it advertised"* (I1), *"is the data contradicting ifself, did the advertisements really take place at the time that they were reported"* (I4). The journalists are interested in relating these findings to their background knowledge or external data sources. All interviewees stated that they would organize nodes into groups (e.g., by political party, by media holding companies). One journalist (I2) also suggested an automatic grouping function.

3.3. Design requirements

Based on these results, we identified seven key requirements (RQ):

RQ1 Ease of use: The interviewees emphasized that tools for data journalists need to be easy to use. Traditional chart types such as bar charts are preferred. More advanced visualization and interaction concepts need to be either self-explanatory or explained thoroughly.

RQ2 Managing large numbers of nodes: A data analysis tool such as *netflower* would typically be needed for networks of non-trivial scale, commonly with 200 to 3,000 nodes. In order to provide an overview, it is necessary to either filter the data or to form meaningful groups.

RQ3 Development over time: The discussed datasets are often collected on a regular basis (i.e., quarterly) and the interviewees expressed particular interest in this temporal aspect. They would investigate patterns or anomalies in the numbers and weights of flows over time in correlation with their knowledge of events in the respective period. They also indicated that comparisons between consecutive quarters might not be as meaningful because seasonal fluctuations could obscure long-term trends.

RQ4 Simple data import: Data journalists need to be capable of analyzing their own data with *netflower*, which is why the preparation of data should not constitute a hurdle. For this reason, a widely used format that can be edited with standard office software (e.g., spreadsheet) should be preferred over specialized database or graph analysis solutions.

RQ5 Export of results: Data journalists would not use their analysis tool to produce charts for their audience because media companies have standardized tools and corporate styles for this purpose. For example, one interviewee suggested that an interface to Datawrappper [dat] could be provided. To facilitate this workflow, *netflower* should allow the export of filtered and annotated data to widely used formats.

RQ6 Analytic provenance: All interviewees expressed the need to take notes and make annotations. However, they largely use special note-taking solutions such as Evernote [eve] or an analog booklet. The sharing of notes was required by only one interviewee (I1), while two (I2, I3) stated that it was not required due to small team sizes. Two interviewees suggested a history mechanism that documents the research process and allows users to revisit points in research.

RQ7 Data protection: Journalists are concerned with keeping their data analysis confidential until they publish their story. Thus, a solution should store the state of analysis locally - and not on a server.

On the whole, four of the seven identified key requirements (RQ1–3, 6) relate directly to visualization and interaction design. More precisely, RQ1 suggests the selection of familiar visual metaphors, RQ2 relates to visual scalability, RQ3 is about time-oriented visualization and interaction capabilities, and RQ6 calls for integrated provenance. RQ4 and RQ5 emphasize the integration into the journalistic workflow, and RQ7 highlights an important prerequisite to ensure user acceptance.

4. Conceptual Design

Based on the requirements described above, we designed *netflower* and iteratively refined it based on the feedback and suggestions of journalists during a series of workshops. Here, we present the main design decisions that led to *netflower's* final design which was also used in the fourth workshop.

Color concept: To keep the interface simple and straightforward (RQ1), we chose a clear and reduced visual design and used only two main colors for the whole application. Green is used for all user interface elements, while pink is used to highlight the important results and data elements in the visualization interface.

Onboarding: To further reduce the learning curve (RQ1), it became one of our core goals – in the course of the workshops – to make *netflower* as self-explanatory as possible. Based on feedback from the workshops, we decided to include a variety of onboarding content, including introductory video material, screenshots and descriptive text. These different media address the onboarding needs of journalists.

Data import: A dedicated data loading screen allows users to choose a file from the file system, to retrieve it from the web via URL, or to select a sample dataset. Even though *netflower* works with network data internally, journalists can load their data into a table in the comma-separated values (CSV) format. Easy inclusion of their own data is a critical step in the journalists' workflow (RQ4). Each row of the table corresponds to one edge of a category between a source node and a target node at a time step with a given weight. In the MTD, for example, one row contains the flow of the "advertising" category from "Stadt Wien" to "Kronen Zeitung" in Q2/2018. While this format contains redundancy, it allows journalists to prepare their own data using standard spreadsheet tools (RQ4) in an easy-to-use fashion (RQ1). The metadata specified in the table headings are also used in the user interface (e.g., to label target nodes as "media institution"). Explanations and data templates are provided to further increase the ease of use (RQ1). Another crucial aspect that is realized here is the protection of sensible data. All loaded data is only stored locally in each client's browser database (RQ7).

Visualization design: The main visualization uses a Sankey diagram to show the time-dependent directed network with weighted edges, ordered by edge weights (see Figure 1). Sankey diagrams are typically used to visualize flows of energy or material throughout processes [RHF05] and have also been used for representing networks (e.g., [RB10, SGW18]). Following various discussions with journalists and visualization experts, we nevertheless decided to use Sankey diagrams for this purpose as they nicely balance usability and expressiveness for our target audience. Sankey diagrams combine the more regular layout of matrix-based representations that includes positional encoding with the more explicit representation of vertices and edges of node-link diagrams, while avoiding color as a less suitable visual variable for edge weights. With this more intuitive setup, Sankey diagrams seem to be more suitable for casual users as well (RQ1).

In Figure 1, we present an example of the Austrian media transparency dataset showing money flows between legal entities and media institutions. We make a small amount of the data available at first glance, so as not to overwhelm the user (RQ2). Therefore, a default space on the screen for the Sankey diagram is defined where the flow can be visualized efficiently. Another reason to reduce the amount of data is related to performance. We encoded information using both hidden and visible data elements and hatched areas in juxtaposition with the nodes of the Sankey diagram (see Figure 3). The area of the hatched rectangle indicates the amount of hidden elements. When hovering over the node, a tooltip appears showing

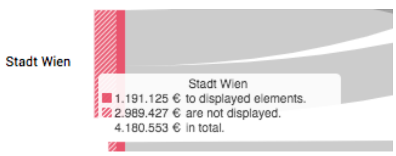

Figure 3: *A node's area encodes the total weights of all ingoing or outgoing edges. The hatched area represents the portion of flows not currently displayed on the screen.*

the amount of money currently visible on the screen, the amount currently not shown, and the total amount of outgoing money. It is possible to load more elements on demand by clicking "Show More" or "Show Less" (see Figure 1 (g)).

Based on the given data structure of dynamic, directed, and weighted network data, we integrated a combination of different common visualization techniques to cover the temporal aspect (RQ3). First, the temporal evolution of one particular flow is displayed after clicking its Sankey edge using a bar chart in a popup window (see Figure 1 (h)). Second, we use bar chart sparklines to represent the amount of outgoing and incoming money over time per node on both sides of the Sankey diagram. To avoid cluttering the interface, textual annotations are displayed only on demand. When hovering over a sparkline, a legend as well as axis descriptions are shown.

Filtering, sorting, ordering: Journalists often have concrete hypotheses regarding the data they want to start working with. Therefore, it is critical to filter, sort and order the data. By default, these operations work globally and influence the entire visualization view (see Figure 1). In particular, *netflower* supports: (i) filtering data in time (RQ3) and by the categories of connection, provided that such a category attribute is in the loaded dataset (Figure 1 (d)), (ii) sorting the data by total amount of source, target and flow as well as ordering it in either ascending or descending order (d), (iii) limiting the data elements by applying value filters to the source, flow and target (RQ2) (e), (iv) searching for a particular data element in the source and also in the target by using the search box (e), and (v) selecting a particular node to have a closer look on it in filtered form. Furthermore, exporting and documentation options are provided. Users can export the data of the currently visible area, which will create a .csv file with the current, filtered dataset. Alternatively, the complete dataset can be exported as a whole. Moreover, the user can create a scalable vector graphics (SVG) file from the current view in order to use the visualization directly (RQ5).

Tags: We found that journalists sometimes have certain ways of grouping the data in mind. To support this task, users can use tagging to organize and group both source and target nodes in *netflower*. Tags can be added in the source file or by adding them directly in the interface by clicking the tag icon next to each node (see Figure 1 (i)). With the tag filter below the search bar, users can reduce the dataset to predefined or self-defined tags (single or multiple). Moreover, an aggregated view can be achieved by switching from a display of nodes to a display of flows between tags (RQ2).

Notebook: To meet requirement RQ6 (analytic provenance), we integrated a notebook that allows for tracking insights and exploration paths. We expected this to be particularly important as journalists would often get interrupted in their workflow and need to be able to easily resume any analytic processes. Furthermore, reproducibility is imperative as a pillar of guaranteeing journalistic integrity. In the interface, the notebook is integrated into the sidebar (Figure 1 (f)). To comply with RQ7, the notebook is implemented on the client side and saved in the local storage. The notes can also be saved as a .txt file and can be re-uploaded in the notebook.

5. Implementation

Netflower is a client-side based web application (http://netflower.fhstp.ac.at/) and does not rely on a server to process or store data, which ensures local storage of the data (RQ7). In order to prevent data loss upon refreshing or navigation and to make the data persistent when closing the browser or turning off the computer, we use localStorage [loc] and IndexedDB [ind] which are included in a wide range of modern browsers. This allows the application to access a database that is related to the main storage and the physical storage on each user's computer, providing not only the local but also the temporal flexibility needed by many journalists. So far, we specifically target the Chrome browser as it has the best performance and allows for all data to be processed on the client side.

Netflower is designed using the Phovea Framework [pho], an open-source visual analysis platform developed as part of the Caleydo project [cal]. Netflower requires only the pure client side of the Phovea Framework as well as TypeScript [typ] with D3.js [BOH11] to render the visualizations. In addition, several smaller extensions or utility libraries are used, such as Bootstrap [boo] for consistent styling and sankey.js [san] for enhancing the default D3 library and creating the main visualization.

6. Validation and results

From 2017 to 2018, we organized four empowerment workshops as part of a larger program that seeks to further educate Austrian journalists (see Figure 2). Our primary goal was to evaluate the *netflower* prototype's usability based on a qualitative user study. Furthermore, two diary studies were conducted in 2018. The following subsections describe the methods we used as well as their results.

6.1. Empowerment workshops

The main underlying purpose of the four workshops was to teach journalists basic DDJ skills, using *netflower* as an example of a data analysis tool. This enabled us to gain valuable feedback about the design of the prototype, while at the same time imparting skills regarding data journalism such as data analysis and visualization. In addition to *netflower*, our workshops also included other aspects of the larger analysis pipeline, such as Microsoft Excel or Google Spreadsheet to clean data, and Datawrapper [dat] as well as Highcharts [hig] to present results. The workshop participants were divided into two groups. One group worked with the *netflower* prototype and the other one with off-the-shelf tools. Participants were free to choose which group they wanted to join.

6.1.1. Method and participants

The workshops were made up of a theoretical and a practical part. During the theoretical part, the participants received an introductory overview of data journalism, data analysis, and visualization in general. In the second, practical part, participants engaged in data analysis and visualization. At the beginning of the practical part, we introduced *netflower* and explained its features. The use of Microsoft Excel/Google Spreadsheet was also briefly demonstrated. For each of the workshops, we provided a number of datasets (e.g., Austrian media transparency dataset [RTR18], UNHCR migration data [UNH18], farm subsidies [Agr18], and OECD aid development [OEC18]).

Following the practical part of using the *netflower* prototype for exploring the data, the participants were asked to complete a questionnaire. A System Usability Scale (SUS) [Bro13] and qualitative questions were used to gain feedback. Additionally, handwritten notes were taken during the observation to collect feedback and problems in the use of *netflower*. The prototype was gradually adapted between the individual workshops based on the feedback and problems identified in the course of these.

Workshop 1 – pilot test: Workshop 1 served as a pilot test of the study and was conducted with eight journalism students (4 male, 4 female). Two of the eight (1 female, 1 male) participated in our study and used version one (v1) of the *netflower* prototype for data analysis.

Workshop 2: Eight (female) persons participated in the second workshop, and again three of them used *netflower* prototype v1. The participants came from the fields of public relations (PR), journalism, and PR consulting. In workshop 2, the prototype (v1) was still at an early stage and its features were limited (no sorting, ordering, grouping, or exporting).

Workshop 3: A total of eleven persons participated in the third workshop and six of them (3 male, 3 female) were part of our study for *netflower* (a managing editor, three journalists, a consultant in the field of digital marketing, a business consultant). In this workshop, the adapted prototype version two (v2) was tested.

Workshop 4: Three journalists (1 female, 2 male) participated in the last workshop and all of them tested *netflower*. In this context, we adapted the concept to create a more practical workshop with only a short introduction session to DDJ. The aim of this workshop was to gain feedback on the final version of *netflower*and to identify new datasets and use cases for dynamic network data. We deemed the latter important in order to learn about the transferability of *netflower* to other journalistic questions, pinpointing *netflower*'s value beyond the datasets we focused on.

6.1.2. Results

Above all, the findings of the workshops helped us to gradually improve the *netflower* prototype. We report on the qualitative feedback in the next paragraphs. All quotes used in the paper have been translated from German by the authors.

Our qualitative analysis indicated that the visual presentation of the data makes the exploration of dynamic network data easier as compared to the analysis in tools such as Google Spreadsheet or Microsoft Excel. For instance, participants in workshop 2 commented: *"You start with an assumption in the netflower tool and you can see if it can be verified – visually. Without the visual presentation, you start to extract the data from the .csv file and then you may realize that you are on the wrong track." (P1)* Another participant emphasized that *netflower "gives an overview of the data with visual help"* (P2). Another person said: *"The tool can also be used for fact-checking and for generating new perspectives and claims to a given topic."(P3)* In workshops 3 and 4, the participants also highlighted the advantage of *netflower* to see the data visually and therefore gain first insights more easily and quickly. On the whole, the observations showed that it was easy for participants to correctly read the Sankey diagram and bar chart sparklines. This finding was interesting for us as it helped us to better understand the level of visual literacy of our target audience.

Sorting & ordering: Prototype v1 did not provide sorting or filtering for nodes or flows, e.g., sorting by source, target nodes or flows, as well as ordering them either ascendingly or descendingly. During workshop 2, participants strongly expressed the need for these functions. E.g., one participant stressed: *"I would like to sort the legal entities and media as well as the money flows, so that I can see the largest money flow between institutions."* Based on the feedback of the workshop and the knowledge of visualization literature [Mun14] regarding the importance of filtering, sorting and ordering, we included different filter and sorting options in v2.

Tagging: Additionally, participants in workshop 2 claimed: *"I would like to group nodes, for example grouping media such as 'Krone, derStandard and Kurier' to a group of daily newspapers."* As a result of these user needs, we enhanced the *netflower* prototype to include the function of grouping nodes via tags (see Figure 1 (i)).

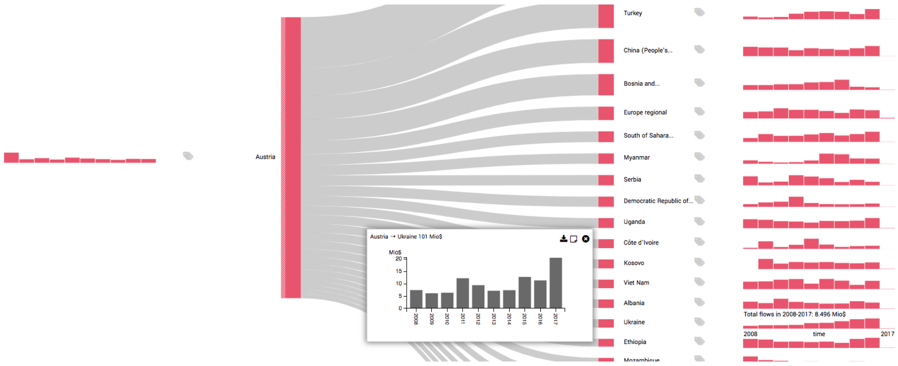

Figure 4: *This figure presents a screenshot of one of the journalist diary studies showing the OECD aid development dataset [OEC18] in netflower. Our journalist participant investigated money flows originating from Austria to receiving countries (on the right of the visualization). The journalist opened the details of the flow Austria-Ukraine, thus learning that the money flow increased over time. The sparklines on the left and right show overall money flows to or from the respective countries over time and are not restricted to the current selection.*

Notebook: Based on the feedback gathered during the problem characterization (see Sec. 3), the interviewees highlighted the need for a notebook where interesting facts and ideas can be documented during data exploration. An unexpected result was that although the notebook was added, none of the participants in all of the four workshops made use of this integrated notebook.

Export: In workshops 2 and 3, we observed that participants had problems creating datasets based on their acquired insights and were confused about how to use these results for presentation tools such as Datawrapper [dat] or Highcharts [hig]. For this reason, we integrated export functions for the full dataset with altered tags in the next version of our prototype (v2).

In workshop 3, participants articulated the need to select a single node in the Sankey diagram to focus on a particular entity and go into more detail in the analysis. This function has been included in the final version of the tool. In workshop 4, we observed that this function was used quite often. In this workshop, we had no further feature requests and participants reported only a limited number of minor issues regarding headings and descriptions. We saw this as a sign of stability in our design process, meaning that *netflower* had reached a level of usability that made it fit for unguided use by journalists in the wild.

Based on the answers of the questionnaire, participants described their knowledge in the use of Microsoft Excel or Google Spreadsheet as good (9 of 12 participants). This self-assessment was, however, in contrast to our observations in the workshop which revealed a lot of difficulties in using MS Excel and Google Spreadsheet. Based on our qualitative observations, we consider the analysis tool literacy of our target audience to be rather low.

Take-home messages:

- Due to its visual representation of the data, *netflower* is easier to use than analysis tools such as Google Spreadsheet or Microsoft Excel.
- The integrated notebook was not used.
- The data export function is crucial for supporting the journalists' workflow from exploration to presentation (Figure 5).

6.2. Diary studies

On the basis of the prototype (v2), two diary studies [LFH10] were conducted in July and August 2018. The following subsection describes the procedure and results.

6.2.1. Method and participants

Two diary studies [LFH10] were conducted with two female journalists (J1–2), both experienced in DDJ. Both journalists were under 30 years of age and each of them worked for a large newspaper. Both have experience in this field based on their studies of more than five years. In their daily work, they focus on Content Management and Feuilleton as well as on implementing data journalistic projects in various fields. In order to give them concrete access points, we provided a variety of datasets for potential analyses: the Austrian media transparency dataset [RTR18], UNHCR migration data [UNH18], farm subsidies [Agr18], and OECD data of aid development [OEC18]. Both journalists decided to use the OECD aid development (2008–2017) dataset for their exploration with *netflower* because of the relevance of the topic. These data include Official Development Assistance (ODA) commitments by EU countries covering data of 2008–2017.

We instructed the participants to explore the data with *netflower* in order to identify interesting aspects which they might write a story about. We aimed to keep the setup of the study as realistic as possible. To achieve this, we gave the participants no personal introduction to the tool. Instead, we wanted to see whether the onboarding strategies we developed based on the workshop feedback would be sufficient to get started. In particular, these strategies explained all the features with screenshots, textual descriptions and videos. All of them were offered on *netflower*'s landing page. In case the participants required assistance, however, they were able to contact us by e-mail or telephone at any time. During the interaction with the tool, our subjects were asked to document their experiences and overall feedback in a *diary*. The participants then sent their feedback via e-mail in the form of textual descriptions and, in some cases, screenshots.

6.2.2. Results

The main feedback put forward by both journalists was that the tool was simple, self-explanatory and very easy to use. The filtering and sorting options were highlighted as helpful for exploration and the participants said that *netflower* provided all the features that they needed. We saw this as an indication that our workshop-based, iterative design approach worked well.

In the diary, one of the journalists documented the following comments regarding the visualization techniques used and information shown: *"Pretty cool, actually, so you can compare how much one country spends on another compared to what it spends on development aid overall."* (J1). Figure 4 shows an exemplary step in the exploration of this journalist.

The other journalist (J2) stated that continuous work with the tool was pretty self-explanatory for most questions that arose in the course of the work. Furthermore, this journalist highlighted the user-friendliness of the tool: *"The explanations on how to use filters were easy to understand. In general, you are guided well through the tool – from data import to the visualization itself and the export of the data"* (J2). The journalists emphasized the importance of exporting data to use these for the presentation of their insights: *"The export also worked well, i.e., if I now decided on a concrete story, I could easily drag it into a visualization tool and then create a graphic"* (J1). Without this feature, the workflow of the data journalist would be interrupted and therefore the *netflower* tool might be not be used.

The journalists stated that they worked with *netflower* for approximately two hours. However, this time span was interrupted again and again, as they could only spend a limited amount of time due to time restrictions in the newsroom.

A particularly surprising finding was that none of the participants actually used our onboarding strategies. They never watched the videos or made use of the introductory descriptions we provided, especially for first-time users. What is more, the subjects did not ask for help during the diary studies. On the one hand, this further supports the thesis that *netflower* is usable indeed. On the other hand, however, it indicates that the simple onboarding techniques we used might not be effective measures to get users started. Although strongly requested in our workshops, this finding would be in line with the large body of literature [BCM86] proving that manuals are mostly ignored.

Take-home messages:

- The *netflower* interface is self-explanatory.
- The participants never used the integrated onboarding strategies.
- The export of data was emphasized as important for the presentation of the insights.

7. Related work

We review related work from two angles. From our application perspective, we discuss current visual analysis tools in the context of DDJ. From an abstract angle, we discuss how the state of the art in the visualization of dynamic networks relates to our work.

Data-driven journalism: Many recent examples of data-intensive newswork [AGO*17] such as the Panama Papers [ICI17a] and the Paradise Papers [ICI17b] have underlined the importance of integrating elements of data science into the journalistic workflow. Weber and Rall analyze the change in the production of news through the advent of visualization methods which are not traditionally part of a journalist's toolbox [WR12]. The importance of visualizations in journalistic works was already addressed in the past, e.g. visualization for storytelling was emphasized by Kosara and Mackinlay [KM13] and Lee et al. [LHRIC15]. Segel and Heer [SH10] explored 58 narrative visualizations from online media. Hullman et al. [HD11] systematically investigated how visualization techniques can be used as a rhetoric device. Satyanarayan and Heer [SH14a] deal with the connection of data visualization and the journalistic necessity of storytelling by developing the Ellipsis model which was designed and tested with the respective user group. Many visualization design environments have been developed to support data journalists in storytelling such as Lyra [SH14b], TimeLineCurator [FBM15], and Timeline Storyteller [BLR*17]. However, these tools use visualization for the presentation of results, while our work focuses on data exploration for story identification.

Beyond the scope of presentation, it is, in fact, becoming more and more common to use data analysis for the identification of topics and the number of available analysis tools for journalists is growing (e.g., OpenRefine [VW13], Trifacta Wrangler [KPHH11], Breve [CJC18], Mirador [Col16], and wtfCSV [DB16]). In many cases, journalists with advanced knowledge and skills in working with data science are the exception rather than the norm [Bra11]. To make data science more easily accessible for journalists in a more general sense, such tools need to utilize simple and easy-to-understand visual interfaces that cater to the exploratory nature of the journalists' activities, for example to allow the incorporation of unconfirmed data that can be used for further investigation [KCCW09, UK15]. A prominent example of such a tool is Overview which allows journalists to analyze large bodies of documents [BISM14]. Beyond tables and texts, however, tool support for journalists is scarce. Despite the increasingly complex and heterogeneous data that journalists face, such as dynamic network data, we are aware of only two enterprise-level platforms for visual analytics for this type of data, Influent [JLGW14] and Linkurious [lin]. As far as we know, there are no dedicated visualization tools that could be readily applied by the participants of our workshops.

Dynamic networks: Our work focuses on visualizing dynamic networks with quantitative flows over time. Many visualization techniques have demonstrated the general value of using interactive visualization to explore time-oriented data [AMST11] and network data [BBDW16, HSS15, KPW14, vLKS*11]. For example, van Ham and Perer [vHP09] use full-text search and a degree-of-interest function to show the relevant parts of a large network in a node-link representation. In contrast, RelEx [SFMB12] is based on a matrix representation of dense directed networks for automotive engineering. Sankey diagrams are often used in engineering to display the flow in a directed network [RHF05]. In social network analysis, Sankey diagrams often depict the development of network communities over time [RB10, RTJ*11, MJF15, VBAW15].

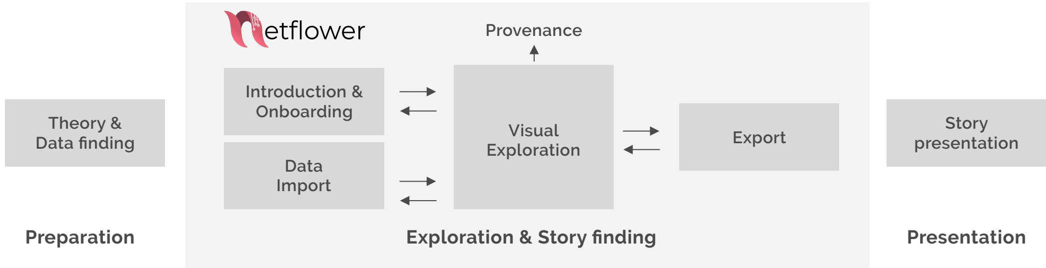

Figure 5: *It takes many different steps to write a data-driven story and present the results to an audience. Starting out by finding a topic, then collecting the data sources, data cleaning (preparation) and analysis (exploration & story finding) as well as building, visualizing and publishing (presentation) the product [UK15]. The netflower prototype focuses on supporting journalists in the exploration and story-finding phase of their work.*

These approaches visually encode how nodes 'flow' between communities between two time points, whereas *netflower* displays edge weights (e.g., money spent) during a time interval and their changes over time. Graph Comics [BKH*16] provide a vocabulary to communicate dynamic networks. While these works provide interesting discussions on visualizing dynamic networks in general, our work specifically focuses on the role of time-dependent quantitative flows in such networks.

The combination of time and quantitative flows in large networks is still an open challenge for visualization research [NAR15]. Only a few visualization designs exist that tackle this problem: The TimeArcTree [GBD09] technique visualizes dynamic directed weighted networks and allows for the aggregation of nodes based on their hierarchy. Visual Adjacency List [HBW14] is a compact network visualization technique that encodes topology by color and also supports dynamic directed weighted networks. EgoSlider [WPZ*15] and egoLines [ZGC*16] are two approaches for ego-centric social network analysis based on dynamic networks that can also have weights and direction. DOSA [vdEvW14] is an approach for the exploration of large networks with multivariate data on the nodes and edges. In this approach, the analyst needs to start with all nodes visible as a point cloud and interactively defines selections in order to get an overview. While these works focus primarily on the visualization techniques themselves, respectively the application for ego-centric social network analysis, our work contributes a design study in close collaboration with practitioners this area.

In this design study, we worked with the media transparency dataset (MTD, cp. Sec. 1), which is structured as a large bipartite network. Some recent visualization designs also focus on such networks. In essence, these approaches contribute automated clustering of nodes based on network topology. VIBR [CXDR18] uses the minimum description length principle for clustering and visualizes the clusters as an adjacency list. WAOW-Vis [PFH*18] hierarchically clusters the nodes with the multiscale dimensionality-reduction technique HSNE and visualizes the clusters in a node-link diagram. BiCFlows [SGW18] uses biclustering and presents the results in a Sankey diagram.

Considering related work that deals with the same dataset (MTD), we identified three approaches. However, these projects target casual users browsing the open-government dataset and do not cater to journalists' needs in story identification. Since 2013, the open source software project Medientransparenz Austria [SBSV] has been providing an interactive online tool that shows the complete MTD. Rind et al. [RPNA16] built a dashboard that presents the different aspects on one interactive screen. Furthermore, BiCFlows [SGW18] conducted their user study on the MTD without its temporal dimension. Apart from that, there is some press coverage on the data and a couple of articles are accompanied by interactive web infographics (e.g., derStandard.at [Ham15], Paroli Magazin [Lan13]).

In short, while some research has been carried out on flows in large dynamic networks, no design studies have been found that target data journalists working with such data. The existing visualization approaches are either too limited, too complex, or tailored to different tasks, thus leaving many of the journalists' needs unaddressed.

8. Reflection, limitations, and future work

The results of the workshops and the diary studies showed that the visual representation of the data can support journalists in visual exploration and story-finding (see Figure 5). The participants emphasized the benefit of exploring the data in a visual form instead of gaining insights from the raw data using table-based tools such as Microsoft Excel or Google Spreadsheet. In our qualitative study, we focused on the usability of *netflower*. For streamlining more detailed design aspects, it would be helpful to add quantitative evaluations with larger numbers of participants and scripted tasks. A comparative study of different tools would be interesting as well. Finally, also even more tightly embedded, long-term case studies [SP06] covering several months of real use in a media outlet would help to assess the domain value of *netflower*.

In our 4-year iterative process of working with journalists, we characterized a common workflow in the way these journalists integrate data into their processes. We built *netflower* to directly support this process and we share it here to allow others to use it as a framework for other journalistic visualization projects. Figure 5 shows the resulting workflow. At the center, we indicate the role that *netflower* plays in this process by supporting users in visual exploration and story-finding. Before the exploration phase starts, the journalists find themselves in a preparation phase during which they search and/or collect data and develop (first) theories with which to challenge the data. We know from the literature in the field of DDJ that the term workflow is generally used when it comes to data-driven journalism [UK15]. Uskali et al. [UK15] identified the following steps: "finding a topic, collecting the data sources, cleaning and analyzing, building & visualizing and publishing the product" [UK15]. The exploratory analysis of large amounts of data with the aim of finding a topic is a central step. The workflow presented in Figure 5 is inspired by the sensemaking loop for intelligence analysis [PC05] and the storytelling process by Lee et al. [LHRIC15]. It provides a first example of a workflow adapted to DDJ.

In workshop 4, we brainstormed journalistic datasets, topics and use cases based on the given data structure of the dynamic network. One of our early concerns with focusing on directed, dynamic, weighted networks was that this focus might be limited to a small set of questions relevant for journalists. In workshop 4, we were very pleasantly surprised that, in a very short amount of time (30 min.), participants were able to identify 27 potential new topics, datasets and use cases that are relevant for them and could be explored with *netflower*. Mentioned topics included traffic streams from country to country, European Song Contest statistics, and investments in rural areas. These questions point to many interesting future directions for studying network visualization in the context of journalism.

Our target user group of journalists and especially those who are interested in data-driven journalism are a small group of experts. Moreover, they are subject to very tight schedules. Thus, 12 journalists taking part in our qualitative studies provides good coverage. We found that the empowerment workshops were an appropriate method to evaluate the *netflower* prototype with journalists. The participants worked intensively with the prototype. Most of the participants spent around 80 minutes with the prototype with some breaks in between. They were very focused on gaining insights and visualizing their findings. Therefore, the documentation of the feedback and problems during the participants' observations and mentoring yields relevant and interesting qualitative results to gradually improve the prototype.

Next we will compare the current state of *netflower* with the requirements defined during problem characterization, thus establishing the paths ahead.

RQ1 Ease of use: Over the duration of the project, the usability of *netflower* reached a good level of maturity. Furthermore, the prototype is accompanied by a clear onboarding guide with text and videos. What we observed in the diary studies with the two journalists was that they never used the onboarding concepts we had created (videos, screenshots and textual descriptions). On the one hand, this indicates that the *netflower* tool is easy to use and self-exploratory for journalists. On the other hand, however, more research is need to explore in depth which onboarding concepts are efficient and why this kind of help is not used at all.

RQ2 Managing large numbers of nodes: By default, *netflower* shows only the latest (time factor) and largest (quantitative factor) flows in the dataset when loading new data. Journalists can choose to display more data, use various filters, or change the sorting and/or ordering in order to refine the visible flows. In addition, they have the option to filter the data by tags that are either added manually or loaded together with the data. While such an approach does not follow the classic "overview first" principle [Shn96], it is similar to the approaches already tested by DOSA [vdEvW14] and by van Ham and Perer [vHP09]. One possible direction for future research is the addition of clustering or community detection methods [CXDR18, PFH*18, RC18, SGW18] as well as the evaluation of these clusters by data journalists.

RQ3 Development over time: *netflower* presents time-oriented aspects as bar charts, both for nodes and for flows. It is also possible to change the time periods displayed in the central Sankey diagram. If the Sankey diagrams of two different periods are to be compared, the journalist would need to open *netflower* in a second browser window and take care that the filters are synchronized. A dedicated comparison mode would be an option for future work.

RQ4 Simple data import: *netflower* loads its data from a single, edge-based CSV file which can be prepared in a spreadsheet program. The loading screen provides detailed information and example files on how the data should be arranged. In the first three workshops, participants worked primarily with the provided sample data. Participants in the fourth workshop had no difficulties preparing their own data for an import into *netflower*.

RQ5 Export of results: *netflower* not only allows the export of flows and individual time series in CSV format, but also the export of the full dataset with custom set tags as well as the generation of a SVG file from the currently visible data flow. This fulfills the design requirement and was appreciated by the test persons.

RQ6 Analytic provenance: In the requirement-gathering phase, most interviewees stated very strongly that they preferred dedicated note-taking tools. To achieve that, *netflower* provides basic notebook features in an expandable sidebar. Furthermore, *netflower* tracks interaction in a log file which could be used to reconstruct the research process. However, it is not possible to bookmark or save a system state to revisit it later. During the workshops and diary studies, the simple provenance functions were not adopted and neither did the participants express the need for more sophisticated functions. *netflower* shares this limitation with a large number of visualization tools. Nevertheless, given the importance of journalists' accountability, an integrated note-taking and history function might be a valuable future addition [SKS*08, XAJK*15].

RQ7 Data protection: *netflower* is built as a web application that communicates only with a server to provide the web page and all resources it needs. The data added by the users is never stored on a server, neither are the notes. This feature was appreciated by participants of the workshops.

Further research should be undertaken to explore how *netflower* can be integrated into the daily routine of data journalists. Therefore, more diary studies as well as instrumentation are needed to evaluate this aspect.We implemented a logging function, capturing the interactions of the participants during their use of *netflower*. However, the logging of data has not yielded any interesting results so far and more work is needed in this respect.

9. Conclusion

We presented a design study of *netflower*, a tool for the visual exploration of dynamic network data in the application domain of journalism. Based on a qualitative user study including empowerment workshops and two diary studies, we validated the visualization design of *netflower*. We learned that the used visualization techniques are appropriate and can support journalists in the exploration and story-finding process. Additionally, we found out that journalists follow a generalizable workflow when working with data-driven storytelling which can be supported with *netflower*. Moreover, journalists in our studies identified a broad range of uses cases that fit the abstraction of flows between nodes in dynamic networks.

Acknowledgements

This work was supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) under the ICT of the Future program via the VALiD project (no. 845598), and by the Austrian Science Fund as part of the VisOnFire project (FWF P27975-NBL).

References

[AGO*17] AUSSERHOFER J., GUTOUNIG R., OPPERMANN M., MATIASEK S., GOLDGRUBER E.: The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism, online first* (2017). doi:10.1177/1464884917700667. 2, 8

[Agr18] AGRARMARKT AUSTRIA: Transparenzdatenbank EU. https://www.transparenzdatenbank.at/, 2018. Accessed: 2018-11-30. 6, 7

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*. Springer, London, 2011. doi:10.1007/978-0-85729-079-3. 2, 8

[BBDW16] BECK F., BURCH M., DIEHL S., WEISKOPF D.: A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum 36*, 1 (2016), 133–159. doi:10.1111/cgf.12791. 8

[BCM86] BLACK J. B., CARROLL J. M., MCGUIGAN S. M.: What kind of minimal instruction manual is the most effective. *SIGCHI Bull. 17*, SI (May 1986), 159–162. doi:10.1145/30851.275623. 8

[BISM14] BREHMER M., INGRAM S., STRAY J., MUNZNER T.: Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Trans. Visualization and Computer Graphics 20*, 12 (2014), 2271–2280. doi:10.1109/TVCG.2014.2346431. 8

[BKH*16] BACH B., KERRACHER N., HALL K. W., CARPENDALE S., KENNEDY J., HENRY RICHE N.: Telling stories about dynamic networks with graph comics. In *Proc. CHI Conf. Human Factors in Computing Systems* (2016), ACM, pp. 3670–3682. doi:10.1145/2858036.2858387. 9

[BLR*17] BREHMER M., LEE B., RICHE N. H., EDGE D., WHITE C., LYTVYNETS K., TITTSWORTH D.: Microsoft Timeline Storyteller. https://timelinestoryteller.com, 2017. Accessed: 2018-12-10. 8

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3: Data-driven documents. *IEEE Trans. Visualization and Computer Graphics 17*, 12 (2011), 2301–2309. doi:10.1109/TVCG.2011.185. 5

[boo] Bootstrap – responsive design. https://getbootstrap.com/. Accessed: 2018-11-22. 5

[Bra11] BRADSHAW P.: Data journalism survey: A mixed picture. European Journalism Centre, http://datadrivenjournalism.net/news_and_analysis/data_journalism_survey_analysis, 2011. Accessed: 2018-12-10. 3, 8

[Bro13] BROOKE J.: SUS: A retrospective. *J. Usability Studies 8*, 2 (2013), 29–40. 3, 6

[cal] Caleydo – research community. https://caleydo.org/. Accessed: 2018-11-22. 5

[CJC18] CESERANI G., JEWETT E., COLEMAN N.: Breve – see your data. http://hdlab.stanford.edu/breve/, 2018. Accessed: 2018-12-10. 8

[Col16] COLUBRI A.: A Mirador for the data forest: From exploration to predictive modeling. European Journalism Centre, http://datadrivenjournalism.net/news_and_analysis/a_mirador_for_the_data_forest_from_exploration_to_predictive_modeling, 2016. Accessed: 2018-12-10. 8

[CXDR18] CHAN G. Y., XU P., DAI Z., REN L.: ViBr: Visualizing bipartite relations at scale with the minimum description length principle. *IEEE Trans. Visualization and Computer Graphics 25*, 1 (2018), 321–330. doi:10.1109/TVCG.2018.2864826. 9, 10

[dat] Datawrapper. https://www.datawrapper.de/. Accessed: 2018-11-16. 4, 6, 7

[DB16] D'IGNAZIO C., BHARGAVA R.: DataBasic: Design principles, tools and activities for data literacy learners. *The Journal of Community Informatics 12*, 3 (2016). 8

[eve] Evernote. https://evernote.com/intl/en/. Accessed: 2018-12-08. 3, 4

[FBM15] FULDA J., BREHMER M., MUNZNER T.: Timelinecurator: Interactive authoring of visual timelines from unstructured text. *IEEE Trans. Visualization and Computer Graphics 22*, 1 (2015), 300–309. doi:10.1109/TVCG.2015.2467531. 8

[GBD09] GREILICH M., BURCH M., DIEHL S.: Visualizing the evolution of compound digraphs with TimeArcTrees. *Computer Graphics Forum 28*, 3 (2009), 975–982. doi:10.1111/j.1467-8659.2009.01451.x. 9

[gep] Gephi. https://gephi.org/. Accessed: 2018-12-07. 2

[Ham15] HAMETNER M.: Inserate: 40,6 Millionen im ersten Quartal. Web-Standard, http://derstandard.at/2000017464403/, 2015. Accessed: 2015-06-15. 9

[HBW14] HLAWATSCH M., BURCH M., WEISKOPF D.: Visual adjacency lists for dynamic graphs. *IEEE Trans. Visualization and Computer Graphics 20*, 11 (2014), 1590–1603. doi:10.1109/TVCG.2014.2322594. 9

[HD11] HULLMAN J., DIAKOPOULOS N.: Visualization rhetoric: Framing effects in narrative visualization. *IEEE Trans. Visualization and Computer Graphics 17*, 12 (2011), 2231–2240. doi:10.1109/TVCG.2011.255. 8

[hig] Highcharts. https://www.highcharts.com/. Accessed: 2018-11-16. 6, 7

[HSS15] HADLAK S., SCHUMANN H., SCHULZ H.-J.: A survey of multi-faceted graph visualization. In *Proc. Eurographics Conf. Visualization – State of The Art Report, EuroVis STAR* (2015), Borgo R., Ganovelli F., Viola I., (Eds.), Eurographics, pp. 1–20. doi:10.2312/eurovisstar.20151109. 8

[ICI17a] ICIJ: The Panama Papers: Exposing the rogue offshore finance industry. https://www.icij.org/investigations/panama-papers/, 2017. Accessed: 2018-12-10. 8

[ICI17b] ICIJ: Paradise Papers: Secrets of the global elite. https://www.icij.org/investigations/paradise-papers/, 2017. Accessed: 2018-12-10. 8

[ind] IndexedDB API – low level API for client side storage of data. https://developer.mozilla.org/en-US/docs/Web/API/IndexedDB_API. Accessed: 2018-11-22. 5

[JLGW14] JONKER D., LANGEVIN S., GAULDIE D., WRIGHT W.: Influent: Scalable transactional flow analysis with entity-relationship graphs. In *Poster Proc. EuroVis* (2014). https://uncharted.software/assets/influent-scalable-transaction-flow-analysis.pdf, Accessed: 2019-02-25. 8

[KCCW09] Kosara R., Cohen S., Cukier J., Wattenberg M.: Panel: Changing the world with visualization. In *Proc. IEEE Visualization* (2009). 8

[KM13] Kosara R., Mackinlay J.: Storytelling: The next step for visualization. *Computer 46*, 5 (2013), 44–50. doi:10.1109/MC.2013.36. 8

[KPHH11] Kandel S., Paepcke A., Hellerstein J., Heer J.: Wrangler: interactive visual specification of data transformation scripts. In *Proc. SIGCHI Conf. Human Factors in Computing Systems, CHI* (2011), ACM, pp. 3363–3372. doi:10.1145/1979442.1979444. 8

[KPW14] Kerren A., Purchase H. C., Ward M. O. (Eds.): *Multivariate Network Visualization*. LNCS 8380. Springer, Cham, 2014. doi:10.1007/978-3-319-06793-3. 8

[Lan13] Lang F.: Medientransparenz – die Zweite. Paroli-Magazin, http://www.paroli-magazin.at/555/, 2013. Accessed: 2013-03-18. 9

[LFH09] Lazar J., Feng J. H., Hochheiser H.: *Research Methods in Human-Computer Interaction*. John Wiley & Sons, Chichester, UK, 2009. 3

[LFH10] Lazar D. J., Feng D. J. H., Hochheiser D. H.: *Research Methods in Human-Computer Interaction*. John Wiley & Sons, 2010. 7

[LHRIC15] Lee B., Henry Riche N., Isenberg P., Carpendale S.: More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications 35*, 5 (2015), 84–90. doi:10.1109/MCG.2015.99. 8, 10

[lin] Linkurious. https://linkurio.us/. Accessed: 2019-02-25. 8

[loc] localStorage – in browser storage facility. https://developer.mozilla.org/en-US/docs/Web/API/Window/localStorage. Accessed: 2018-11-22. 5

[MJF15] Morini M., Jensen P., Flandrin P.: Temporal evolution of communities based on scientometrics data. In *Sciences des données et humanités numeriques* (Paris, France, Nov. 2015). 8

[Mun14] Munzner T.: *Visualization Analysis and Design*. CRC Press, 2014. 6

[NAR15] Niederer C., Aigner W., Rind A.: Survey on visualizing dynamic, weighted, and directed graphs in the context of data-driven journalism. In *Proc. Int. Summer School on Visual Computing* (2015), Schulz H.-J., Urban B., Freiherr von Lukas U., (Eds.), Frauenhofer Verlag, pp. 49–58. 9

[NRA*16] Niederer C., Rind A., Aigner W., Ausserhofer J., Gutounig R., Sedlmaier M.: Visual exploration of media transparency for data journalists: Problem characterization and abstraction. In *Proc. 10th Forschungsforum der österreichischen Fachhochschulen* (2016), FH des BFI Wien. 3

[OEC18] OECD: Detailed aid statistics: ODA commitments. OECD International Development Statistics (database), 2018. Accessed: 2018-11-30. doi:10.1787/data-00068-en. 6, 7

[PC05] Pirolli P., Card S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. Int. Conf. on Intelligence Analysis* (2005). 10

[PFH*18] Pezzotti N., Fekete J.-D., Höllt T., Lelieveldt B. P. F., Eisemann E., Vilanova A.: Multiscale visualization and exploration of large bipartite graphs. *Computer Graphics Forum 37*, 3 (June 2018), 549–560. doi:10.1111/cgf.13441. 9, 10

[pho] Phovea – framework for designing visualizations. https://wiki.datavisyn.io/phovea. Accessed: 2018-11-22. 5

[RB10] Rosvall M., Bergstrom C. T.: Mapping change in large networks. *PLOS ONE 5*, 1 (Jan. 2010), e8694. doi:10.1371/journal.pone.0008694. 5, 8

[RC18] Rossetti G., Cazabet R.: Community discovery in dynamic networks: A survey. *ACM Comput. Surv. 51*, 2 (Feb. 2018). doi:10.1145/3172867. 10

[RHF05] Riehmann P., Hanfler M., Froehlich B.: Interactive Sankey diagrams. In *Proc. IEEE Symp. Information Visualization, INFOVIS* (2005), pp. 233–240. doi:10.1109/INFVIS.2005.1532152. 5, 8

[RPNA16] Rind A., Pfahler D., Niederer C., Aigner W.: Exploring media transparency with multiple views. In *Proc. 9th Forum Media Technology* (2016), Aigner W., Schmiedl G., Blumenstein K., Zeppelzauer M., (Eds.), CEUR-WS.org, pp. 65–73. 2, 9

[RTJ*11] Reda K., Tantipathananandh C., Johnson A., Leigh J., Berger-Wolf T.: Visualizing the evolution of community structures in dynamic social networks. *Computer Graphics Forum 30*, 3 (2011), 1061–1070. doi:10.1111/j.1467-8659.2011.01955.x. 8

[RTR18] RTR GmbH: Katalog Medientransparenz Datenmeldung. Offene Daten Österreich, 2018. https://www.data.gv.at/katalog/dataset/c3a8d866-408e-4042-893e-99710b4db683. Accessed: 2018-11-27. 2, 6, 7

[san] SankeyJS – Enhancement of D3 standard library. https://github.com/d3/d3-sankey. Accessed: 2018-11-22. 5

[SBSV] Salhofer P., Basyouni A., Stibler M., Vrecer S.: Medientransparenz Austria. http://www.medien-transparenz.at/. Accessed: 2016-09-09. 9

[Sch08] Schmidt M.: The Sankey diagram in energy and material flow management: Part I: History. *Journal of Industrial Ecology 12*, 1 (2008), 82–94. doi:10.1111/j.1530-9290.2008.00004.x. 2

[SFMB12] Sedlmair M., Frank A., Munzner T., Butz A.: RelEx: Visualization for actively changing overlay network specifications. *IEEE Trans. Visualization and Computer Graphics 18*, 12 (2012), 2729–2738. doi:10.1109/TVCG.2012.255. 8

[SGW18] Steinböck D., Gröller E., Waldner M.: Casual visual exploration of large bipartite graphs using hierarchical aggregation and filtering. In *Proc. Int. Symp. Big Data Visual and Immersive Analytics, BDVA* (2018), pp. 157–166. doi:10.1109/BDVA.2018.8533894. 5, 9, 10

[SH10] Segel E., Heer J.: Narrative visualization: Telling stories with data. *IEEE Trans. Visualization and Computer Graphics 16*, 6 (2010), 1139–1148. doi:10.1109/TVCG.2010.179. 8

[SH14a] Satyanarayan A., Heer J.: Authoring narrative visualizations with ellipsis. *Computer Graphics Forum 33*, 3 (2014), 361–370. doi:10.1111/cgf.12392. 8

[SH14b] Satyanarayan A., Heer J.: Lyra: An interactive visualization design environment. *Computer Graphics Forum 33*, 3 (2014), 351–360. doi:10.1111/cgf.12391. 8

[Shn96] Shneiderman B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. IEEE Symp. Visual Languages, VL* (1996), pp. 336–343. doi:10.1109/VL.1996.545307. 10

[SKS*08] Scheidegger C., Koop D., Santos E., Vo H., Callahan S., Freire J., Silva C.: Tackling the Provenance Challenge one layer at a time. *Concurrency and Computation: Practice and Experience 20*, 5 (2008), 473–483. doi:10.1002/cpe.1237. 10

[SMM12] Sedlmair M., Meyer M., Munzner T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Visualization and Computer Graphics 18*, 12 (2012), 2431–2440. doi:10.1109/TVCG.2012.213. 2

[SP06] Shneiderman B., Plaisant C.: Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (New York, NY, USA, 2006), BELIV '06, ACM, pp. 1–7. doi:10.1145/1168149.1168158. 9

[tab] Tableau. https://www.tableau.com/. Accessed: 2018-12-07. 2

[typ] TypeScript – superset of plain JavaScript. https://www.typescriptlang.org/. Accessed: 2018-11-22. 5

[UK15] USKALI T. I., KUUTTI H.: Models and streams of data journalism. *The Journal of Media Innovations 2*, 1 (2015), 77–88. doi:10.5617/jmi.v2i1.882. 8, 9, 10

[UNH18] UNHCR: UNHCR Statistics – The World in Numbers. http://popstats.unhcr.org/, 2018. Accessed: 2018-11-13. 6, 7

[VBAW15] VEHLOW C., BECK F., AUWÄRTER P., WEISKOPF D.: Visualizing the evolution of communities in dynamic graphs. *Computer Graphics Forum 34*, 1 (2015), 277–288. doi:10.1111/cgf.12512. 8

[vdEvW14] VAN DEN ELZEN S., VAN WIJK J. J.: Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *IEEE Trans. Visualization and Computer Graphics 20*, 12 (2014), 2310–2319. doi:10.1109/TVCG.2014.2346441. 9, 10

[vHP09] VAN HAM F., PERER A.: "search, show context, expand on demand": Supporting large graph exploration with degree-of-interest. *IEEE Trans. Visualization and Computer Graphics 15*, 6 (Nov. 2009), 953–960. doi:10.1109/TVCG.2009.108. 8, 10

[vLKS*11] VON LANDESBERGER T., KUIJPER A., SCHRECK T., KOHLHAMMER J., VAN WIJK J. J., FEKETE J.-D., FELLNER D. W.: Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum 30*, 6 (2011), 1719–1749. doi:10.1111/j.1467-8659.2011.01898.x. 2, 8

[VW13] VERBORGH R., WILDE M. D.: *Using OpenRefine*. Packt, 2013. 8

[WPZ*15] WU Y., PITIPORNVIVAT N., ZHAO J., YANG S., HUANG G., QU H.: egoSlider: Visual analysis of egocentric network evolution. *IEEE Trans. Visualization and Computer Graphics 22*, 1 (2015), 260–269. doi:10.1109/TVCG.2015.2468151. 9

[WR12] WEBER W., RALL H.: Data visualization in online journalism and its implications for the production process. In *Proc. 16th Int. Conf. Information Visualisation, IV* (2012), IEEE, pp. 349–356. doi:10.1109/IV.2012.65. 8

[XAJK*15] XU K., ATTFIELD S., JANKUN-KELLY T. J., WHEAT A., NGUYEN P. H., SELVARAJ N.: Analytic provenance for sensemaking: A research agenda. *IEEE Computer Graphics and Applications 35*, 3 (2015), 56–64. doi:10.1109/MCG.2015.50. 10

[ZGC*16] ZHAO J., GLUECK M., CHEVALIER F., WU Y., KHAN A.: Egocentric analysis of dynamic networks with EgoLines. In *Proc. CHI Conf. Human Factors in Computing Systems* (2016), ACM, pp. 5003–5014. doi:10.1145/2858036.2858488. 9