

PubViz: Lightweight Visual Presentation of Publication Data

A. Rind, A. Haberson, K. Blumenstein, C. Niederer, M. Wagner, and W. Aigner

Institute of Creative\Media\Technologies, St. Pölten University of Applied Sciences, Austria

Abstract

Publications play a central role in presenting the outcome of scientific research but are typically presented as textual lists, whereas related work in visualization of publication focuses on exploration – not presentation. To bridge this gap, we conducted a design study of an interactive visual representation of publication data in a BibTeX file. This paper reports our domain and problem characterization as well as our visualization design decisions in light of our user-centered design process including interviews, two user studies with a paper prototype and a d3.js prototype, and practical application at our group’s website.

Categories and Subject Descriptors (according to ACM CCS): Information Interfaces and Presentation [H.5.2]: User Interfaces—Graphical user interfaces

1. Introduction

One’s publications are probably the most vital assets for early stage and senior researchers alike. Despite their central role as scientific track record, publication lists of a single researcher or research groups are still commonly represented as textual lists. These can, for example, be grouped by the type of publication or sorted by year of publication. However, the representation itself as well as the level of interactivity for online versions are rather limited.

This paper is a design study [SMM12] that presents the domain and problem characterization, conceptual design, and implementation of an interactive visual representation of publication data called PubViz. The designed representations are interactive and coordinated, focusing on different aspects such as development over time, publication type distribution, co-authors, and keywords. Input data is given in the established BibTeX format [Pat10] and the interactive views are implemented based on standard web technologies such as d3.js [BOH11] in order to be easily embeddable into webpages. PubViz is available as free and open source software on GitHub [HC17].

The main contributions of the paper at hand are 1) a systematic domain and problem characterization manifested in concrete personas and scenarios that may be used for further work in the area, 2) a specific visualization and interaction design along with 3) empirical evidence collected in the course of a user-centered design process and 4) an open and reusable implementation in JavaScript.

2. Design Method and Evolution

For the design of PubViz, we used a four-level user-centered design process [RSP15], iteratively involving persons from the target audience throughout the process.

In an initial phase, we conducted interviews with six university employees to specify the context of use and to derive requirements. The interviews were conducted at their workspace to gain a deeper insight into their natural working environment. The main aims of these interviews were to estimate the general need for the visualization of publication data, find essential expectations on such a tool, and identify the most important data in a publication list. The results of the first stage are presented in more detail in Section 3.

Based on analysis of the outcome of the conducted interviews, representative BibTeX data, and related work (Section 4), we sketched possible visualizations of publication data (Figure 1). In this early development stage, we performed a formative usability evaluation [TQ05] of a Wizard of Oz [Nie93] setting with six subjects to find out which visualizations are most suitable (Figure 2).

Following the formative user study a high-fidelity prototype was developed and validated in the form of a usability study [Nie93] with four test persons of the target group. The main goal of the usability study was to get feedback from users and notice usability problems previously unknown.

Afterwards, we deployed the software on the website of our research group (Section 6) and have been using it for several months by the time of publication.

3. Users, Tasks, and Data

In our initial problem characterization, we identified three distinct goals for visualization of publications and developed a persona [CRC07] for each:

Autobiographical Presentation: Scientists have a high interest to make their own publications as visible as possible. On the one hand, they want their research results to be found, applied, and

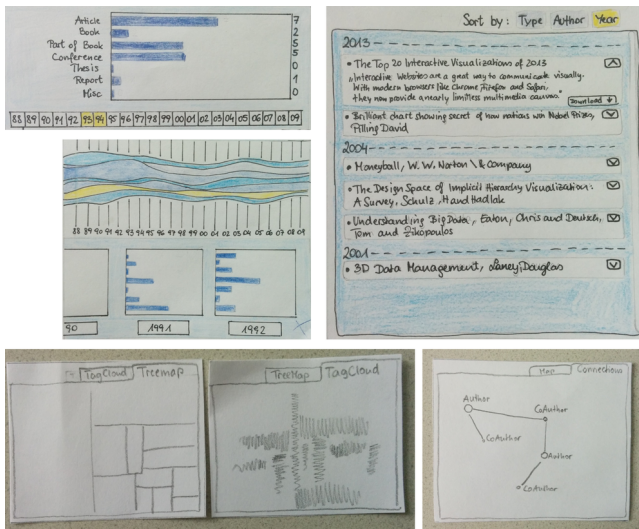


Figure 1: First round of UI and visualization sketches.



Figure 2: Wizard of Oz paper prototype.

cited. On the other hand, their publication record can be a key asset in finding collaboration partners and in gaining funding.

Sigrid Zippelin, the first persona, is a 47-year-old university professor and department head in IT security. On the occasion of giving a conference keynote, she creates an appealing landing page for her research with PubViz. She has her publication data available in Zotero [Roy17] and can export it as BibTeX. Her busy schedule would not allow her to use a more complex tool or develop a professional solution herself.

Presentation by Third Parties: Academic institutions, funding bodies, scientific societies, and many other third parties typically report publication lists prominently to represent their competence and/or scientific impact.

Julia Mischalek, the second persona, is a 32-year-old organizational assistant at a small university, where she is responsible for publication lists on the website and various reports. Since she receives publication data in various formats, she maintains a Zotero database. Seeing the results of her drudgery in PubViz increases her motivation, also the scientific staff is more engaged to use the lists, e.g., to find internal collaborators. Finally, Julia can export charts from PubViz and use them in printed reports.

Exploration: Scientists also collect publication lists and analyze them in order to find previously unknown literature, identify key experts, and get a general overview of a field of research.

Hans Peter Guggel, the third persona, is a 20-year-old bachelor student, who uses PubViz to find a topic and a supervisor for his thesis based on the BibTeX data available from his university.

Overall, the users are scientists, students, or other members of a scientific organization. While the exploration goal is supported by various related work, we identified an unaddressed need for a well-designed but easy-to-setup visual presentation of publication lists.

For a minimal setup effort for users, we decided to use the **BibTeX file format** [Pat10] for input of publication data. With a history of more than 30 years, the BibTeX file format is now a de-facto standard for scientific publications and is widely supported by tools and websites. BibTeX stores publication data in a specific plain text format. One publication item is described by a publication type, a citation key, and key-value pairs for various bibliographic attributes. There are a number of publication types such as article or book. Depending on the publication type, different attributes are required or optionally allowed such as authors, journal, pages, URL, etc.

To determine, which bibliographic attributes should have priority in our visualization, we made episodic observations from analyzing 99 items in a representative BibTeX file and the initial interviews of 6 subjects. Authors, title, and publication year were most important to our 6 subjects and also available in most of the example data. URLs were also frequently available in the data and regarded as rather important. Abstract and keywords were rather important but missing for more than half of the items. Citations had average importance as well, but are not available in BibTeX. Edition, publisher, country of origin, and location were answered as least important. However, our subjects stated that a solution should also contain a common list of the complete bibliographic records for the case that a visitor is unwilling to use other interaction.

It should, however, be noted that various tools use BibTeX with different type or attribute names. We consolidated the different attributes to publication type, year, authors, keywords, title, journal/book title, pages, abstract, URL. Furthermore, we merged publication types to 7 categories: article, book, part of book, conference, thesis, report, misc. This data transformation resulted in a table dataset structure with publications items as rows and bibliographic attributes as columns [Mun14].

4. Related Work

Scientific publications are an interesting data source for visualization research: Not only do they allow for retrospection of the scientific community (e.g., [FGP04, HGEF07, XSSS14, IHK*16,

IIS*17]), but they also offer multiple data facets such as time, text, or undirected networks of co-authors, and directed networks of citations, and sets of keywords [BKW16, FHKM16].

Many solutions focus on citations. For example, VisNavi [SMM13] browse a citation graph starting from a seed paper in a hyperbolic tree. CiteVis [SCH*13] shows a paper collection and applies interactive highlighting for citations. Impact glyphs [MML16] provide a dense representation of a paper's references and citations. PaperQuest [PEM16] supports reading decisions based on citation metrics and previously selected "core" publications. CiteRivers [HHKE16] combines the analysis of citation patterns with topics extracted from the publications' text in a multiple coordinated views. Van Ham [vH04] and Burch et al. [BPW15] presented two solutions to explore publications by keywords and discover topical clusters based on analytical preprocessing. KeyVis [IIS*17] was designed to find related work based on keywords and make informed keyword choices for future publications. ScholarTree [FCM16] uses a botanical metaphor for the dynamic network of a scientist's co-authors. For this it represents sophisticated features like first authorship and tiers of venues. Keshif [YEB16] is a general purpose solution for browsing and filtering of table data that has also been demonstrated with the publication record of Ben Bederson [Bed17]. SurVis [BKW16] is specialized solution to present and analyze scientific literature that was designed for authors of state-of-the-art reports to curate and present their references. It includes keywords, authors, citations, publication clustering, teaser images, and flexible selectors. For more solutions, a recent survey on literature and patent visualization [FHKM16] may be consulted.

Discussion. Existing solutions do not address the presentation goals characterized above. Visualizations like PaperQuest or CiteRivers are designed for users willing to invest time for training and exploration. Other visualizations like CiteVis or ScholarTree can be used for presentation but focus on overview of some aspects but do not include publication list, which our test persons need for unengaged visitors. Even well-designed and usable solutions like Keshif or SurVis might be too overloaded for visitors of an autobiographical presentation. Another issue is setup effort for the user curating the visualization. For example, Keshif requires data as a spreadsheet or with a custom conversion code or SurVis needs the curator to run a Python script for BibTeX preprocessing.

5. Visualization Design

Guided by the interviews and inspired by existing solutions (e.g., [Bed17]) as well as collections of visualizations for presentation purposes, we produced several iterations of sketches, paper prototypes, and finally a functional prototype [KKUW07, p. 50] using d3.js [BOH11]. We collected further user feedback in two rounds of usability tests. Next, we will explain our design decisions based on the final prototype as shown in Figure 3:

Time: A bar chart of the number of works published each year (Figure 3:1) provides a temporal overview. We experimented with various visualization methods [AMST11] such as a stream graph [BW08] of publication types but the Wizard of Oz test

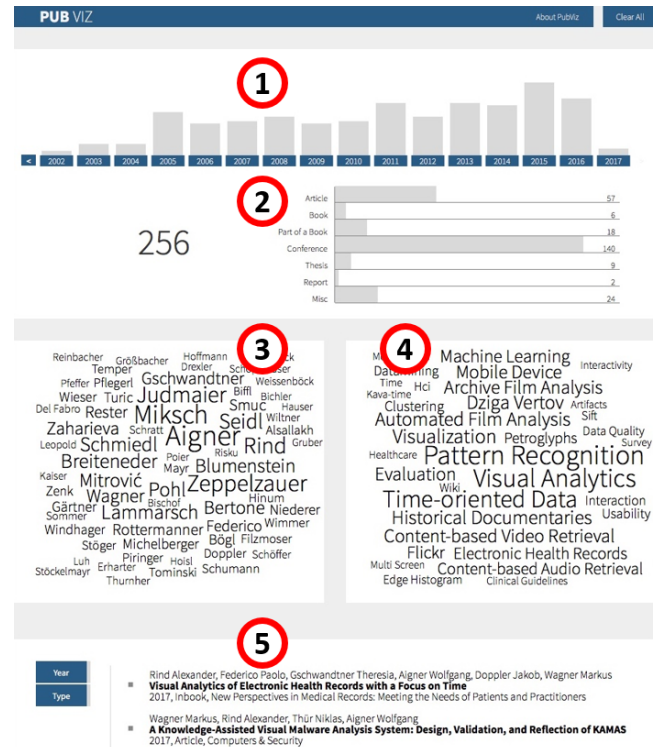


Figure 3: Illustration of the high-fidelity prototype including: 1) temporal bar charts for number of publications by year; 2) on the left the overall number of publications, and on the right a bar chart to break down the count based on publication types; 3) a word cloud of authors; 4) a word cloud of keywords; and 5) the contained publications which can be sorted by year and by type.

revealed that the bar chart was more understandable. During further deployment, we added temporal scrolling, if the covered time span gets too long.

Publication Types: A horizontal bar chart (Figure 3:2) shows how different types of publications compare to each other. We grouped some types together to reduce the space needed. In the early designs, we intended to encode publication type by color hue in various other views. However, this conflicted with the color used for highlighting and we decided that communicating filter state clearly has higher priority. Furthermore, PubViz also shows the total number of publications in a central position (to the left of the publication type bars).

Authors: A word cloud [VW08] (Figure 3:3) shows the names of the publications' authors with a font size based on each author's publication count. In the trade-off between an aesthetically optimized layout and searchability, we settled for a non-sorted layout with horizontal words only that do not change their position on interaction. If the publication list contains too many authors only the top authors are shown.

Initially, we planned a node-link diagram to emphasize co-authorship with edges of different thickness. However, when tested with representative data, this network view was too cluttered – in particular for the limited space available.

Keywords: Another word cloud (Figure 3:4) represents the fre-

quency of the top keywords in the publication list. While the initial interviews did not yield keywords among the most important attributes, the Wizard of Oz and the prototype test revealed that the keyword word cloud was a very useful and well accepted feature by the subjects.

Complete Records: A scrollable list (Figure 3:5) in the bottom of the page shows the bibliographic records of all publications. Visitors can sort the list by year or by type. Each record can be enlarged to show the abstract if present and the given URL is available as a hyperlink. The need for such a list in addition to various overview visualizations was raised by the subjects of the initial interviews after we showed them Ben Bederson's website [Bed17] and confirmed in subsequent usability tests.

A visitor has the ability to select a bar or a word in the overview visualizations (i.e. bar charts and word clouds) for filtering the publication list. In addition all related marks in all views are highlighted [Rob07]. For example, if the user clicks on one year in the temporal bar chart (Figure 4), the bar will be highlighted, bar's label will be displayed with bold font, the list in the bottom will be filtered, the number of matching publications will be displayed in front of the total number, the publication type bar chart will highlight the proportion of work published in that year, and the word clouds will highlight authors and keywords of these filtered publications. It is possible to select more than one mark of a view, which is interpreted as logical OR, and to combine selection in different views, which is interpreted as logical AND [AS94].

The usability of coordinated filtering was a focus of our user studies. Since we cannot expect visualization training from visitors of the publication lists, we settled for relatively weak filtering possibilities. To make the selection state recognizable while maintaining an aesthetic overall appearance, we encode the selection of marks with a bold font only. Furthermore, we added a "clear all" button in the top right corner that resets all selections.

6. Application Examples

To show how PubViz can be used, we demonstrate it for presenting our own publications and exploring a body of literature.

Presenting Own Publications. As a research group or a single researcher, PubViz can be used to present an overview of your publications having an easy way to explore all publications visually. An example is the visualization of publications of a research group [Res17]. The dataset includes 256 items. PubViz is connected to the content management system Drupal via a live BibTeX export which makes it maintainable in the long-term. In addition, the site uses URL filters via GET parameters (e.g., `f[author]`) to prefilter the whole dataset for showing the publications for a specific author.

Exploring Literature. For testing the scalability of PubViz, we used data from the 'vispubdata' project [IHK*16]. PubViz can open the complete dataset and provide an overview of its 2,753 records but interactions are slow on a standard office computer (Intel Core i5-3230M; Chrome 57). It has acceptable reaction times for a prefiltered subset of all 647 publications of the InfoVis conference (from 1995 up to 2015). Here we can recognize in the temporal bar chart that publication count peaked in 2004 (Figure 4) [Rin17]. This

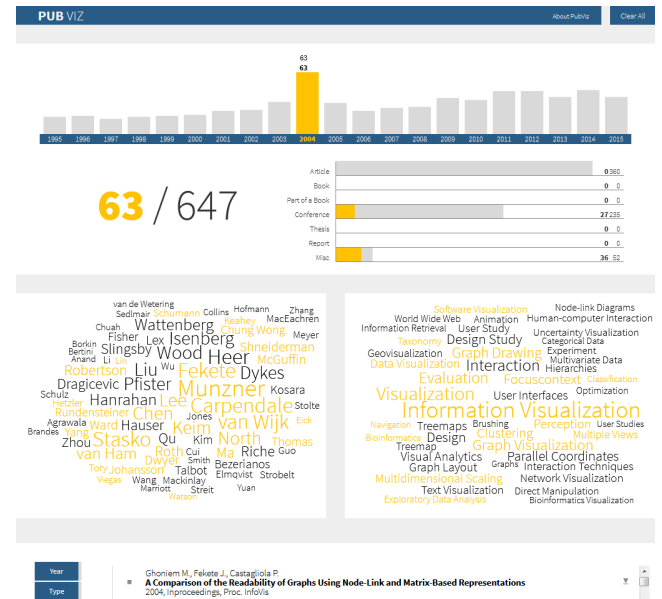


Figure 4: PubViz with all publications of the InfoVis conference from 1995 to 2015. The year 2004 is highlighted.

was the year when posters were also included in the proceedings, which can be seen in the bar chart presentation of the publication types (posters are listed as 'misc'). However, deeper exploration of subsets of records is not possible as the design of PubViz is optimized for presentation to untrained visitors.

7. Conclusions

Scientific researchers want to present their publication record in more accessible way than as a long list of textual references. We conducted a visualization design study to develop a visually appealing and easy-to-setup solution. PubViz takes the established BibTeX format as input, uses the widely known visualization methods bar charts and word clouds, and provides interactive filtering options. Two rounds of user studies with a Wizard of Oz paper prototype and a functional d3.js prototype as well as a long-term application at our research group website serve as validation for our design decisions – in particular the interactive filtering concept did not cause any usability problems. Our long-term application furthermore showed that PubViz works well for a scale of 25 to 250 publications. Thus, it meets the real-world need of autobiographical presentation for individuals, projects, or small research groups.

Supplementary Materials. Research data, software, and reusable figures can be found at <http://phaidra.fhstp.ac.at/o:1933>.

Acknowledgements. We thank M. Fabi, S. Rasul, E. Schnaitt, and P. Theisen from the initial development team; G. Rottermann and N. Thür for technical support; A. Hannauer and P. Federico for literature research; as well as the test persons. This work was partially supported by the Austrian Science Fund (FWF) via the KAVA-Time and VisOnFire projects (no. P25489 and P27975) and the Austrian Ministry for Transport, Innovation and Technology (BMVIT) under the ICT of the future program via the VALiD project (no. 845598).

References

- [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. 3
- [AS94] AHLBERG C., SHNEIDERMAN B.: Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. SIGCHI Conf. Human factors in computing systems* (1994), ACM, pp. 313–317. doi:10.1145/191666.191775. 4
- [Bed17] BEDERSON B.: Ben Bederson's publication, 2017. accessed: 2017-02-26. URL: <https://www.cs.umd.edu/users/bederason/papers/>. 3, 4
- [BKW16] BECK F., KOCH S., WEISKOPF D.: Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Trans. Visualization and Computer Graphics* 22, 1 (Jan. 2016), 180–189. doi:10.1109/TVCG.2015.2467757. 3
- [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. 1, 3
- [BPW15] BURCH M., POMPE D., WEISKOPF D.: An analysis and visualization tool for DBLP data. In *Proc. 19th Int. Conf. Information Visualisation* (2015), pp. 163–170. doi:10.1109/iv.2015.38. 3
- [BW08] BYRON L., WATTENBERG M.: Stacked graphs – geometry & aesthetics. *IEEE Trans. Visualization and Computer Graphics* 14, 6 (Nov. 2008), 1245–1252. doi:10.1109/TVCG.2008.166. 3
- [CRC07] COOPER A., REIMANN R., CRONIN D.: *About Face 3: The Essentials of Interaction Design*. Wiley, Indianapolis, IN, USA, 2007. 1
- [FCM16] FUNG T. L., CHOU J. K., MA K. L.: A design study of personal bibliographic data visualization. In *Proc. IEEE Pacific Visualization Symp. PacificVis* (Apr. 2016), pp. 244–248. doi:10.1109/PACIFICVIS.2016.7465279. 3
- [FGP04] FEKETE J.-D., GRINSTEIN G., PLAISANT C.: IEEE InfoVis 2004 Contest, the history of InfoVis, 2004. accessed: 2017-02-26. URL: <http://www.cs.umd.edu/hcil/iv04contest/>. 2
- [FHKM16] FEDERICO P., HEIMERL F., KOCH S., MIKSCH S.: A survey on visual approaches for analyzing scientific literature and patents. *IEEE Trans. Visualization and Computer Graphics* (2016). Published online 16 Sep 2016. doi:10.1109/TVCG.2016.2610422. 3
- [HC17] HABERSON A., CONTRIBUTORS: PubViz. GitHub repository, 2017. accessed: 2017-02-26. URL: <https://github.com/fhstp/PubViz>. 1
- [HGEF07] HENRY N., GOODSELL H., ELMQVIST N., FEKETE J.-D.: 20 years of four HCI conferences: A visual exploration. *International Journal of Human-Computer Interaction* 23, 3 (Dec. 2007), 239–285. doi:10.1080/10447310701702402. 2
- [HHKE16] HEIMERL F., HAN Q., KOCH S., ERTL T.: CiteRivers: Visual analytics of citation patterns. *IEEE Trans. Visualization and Computer Graphics* 22, 1 (Jan. 2016), 190–199. doi:10.1109/TVCG.2015.2467621. 3
- [IHK*16] ISENBERG P., HEIMERL F., KOCH S., ISENBERG T., XU P., STOLPER C. D., SEDLMIR M., CHEN J., MÖLLER T., STASKO J. T.: vispubdata.org: A metadata collection about IEEE Visualization (VIS) publications. *IEEE Trans. Visualization and Computer Graphics* (2016). Published online 5 Oct 2016. doi:10.1109/TVCG.2016.2615308. 2, 4
- [IIS*17] ISENBERG P., ISENBERG T., SEDLMIR M., CHEN J., MÖLLER T.: Visualization as seen through its research paper keywords. *IEEE Trans. Visualization and Computer Graphics* 23, 1 (Jan. 2017), 771–780. doi:10.1109/TVCG.2016.2598827. 2, 3
- [KKUW07] KULYK O., KOSARA R., URQUIZA J., WASSINK I.: Human-centered aspects. In *Human-Centered Visualization Environments*, Kerren A., Ebert A., Meyer J., (Eds.), LNCS 4417. Springer, Berlin, 2007, pp. 13–75. 3
- [MML16] MAGUIRE E., MONTULL J. M., LOUPPE G.: Visualization of publication impact. In *Proc. EuroVis 2016 – Short Papers* (2016), Bertini E., Elmqvist N., Wischgoll T., (Eds.), Eurographics Association, pp. 103–107. doi:10.2312/eurovisshort.20161169. 3
- [Mun14] MUNZNER T.: *Visualization Analysis and Design*. AK Peters/CRC, Boca Raton, 2014. 2
- [Nie93] NIELSEN J.: *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993. 1
- [Pat10] PATASHNIK O.: bibtex – process bibliographies for LaTeX, etc, 2010. accessed: 2017-02-26. URL: <https://www.ctan.org/pkg/bibtex>. 1, 2
- [PM16] PONSARD A., ESCALONA F., MUNZNER T.: PaperQuest: A visualization tool to support literature review. In *Proc. CHI Conf. Extended Abstracts Human Factors in Computing Systems, CHI EA* (2016), ACM, pp. 2264–2271. doi:10.1145/2851581.2892334. 3
- [Res17] RESEARCH GROUP MEDIA COMPUTING: Publications of Media Computing Research Group, 2017. accessed: 2017-02-26. URL: <http://mc.fhstp.ac.at/pubviz>. 4
- [Rin17] RIND A.: Publications at InfoVis conferences from 1995 to 2015, 2017. accessed: 2017-04-24. URL: <http://alex-rind.github.io/PubViz/?ieee-vis-infovis.bib>. 4
- [Rob07] ROBERTS J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Proc. Conf. Coordinated and Multiple Views in Exploratory Visualization, CMV* (2007), pp. 61–71. doi:10.1109/CMV.2007.20. 4
- [Roy17] ROY ROSENZWEIG CENTER FOR HISTORY AND NEW MEDIA: Zotero, 2017. accessed: 2017-02-26. URL: <https://www.zotero.org/>. 2
- [RSP15] ROGERS Y., SHARP H., PREECE J.: *Interaction Design: Beyond Human-Computer Interaction*, 4 ed. John Wiley & Sons, Chichester, UK, 2015. 1
- [SCH*13] STASKO J., CHOO J., HAN Y., HU M., PILEGGI H., SADANA R., STOLPER C. D.: CiteVis: Exploring conference paper citation data visually. In *Poster Proc. IEEE InfoVis* (2013). 3
- [SMM12] SEDLMIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2431–2440. doi:10.1109/TVCG.2012.213. 1
- [SMM13] SAAD F., MATHIAK B., MUTSCHKE P.: Supporting literature review by searching, visualizing and navigating related papers. In *Proc. Int. Conf. Cloud and Green Computing* (Sept. 2013), pp. 363–368. doi:10.1109/CGC.2013.64. 3
- [TQ05] THEOFANOS M., QUESENBERY W.: Towards the design of effective formative test reports. *J. Usability Studies* 1, 1 (Nov. 2005), 27–45. 1
- [vH04] VAN HAM F.: Case study: Visualizing visualization. In *Proc. IEEE Symp. Information Visualization* (Oct. 2004). doi:10.1109/INFVIS.2004.14. 3
- [VW08] VIÉGAS F. B., WATTENBERG M.: Tag clouds and the case for vernacular visualization. *interactions* 15, 4 (July 2008), 49–52. doi:10.1145/1374489.1374501. 3
- [XSS14] XU P., STOLPER C., SAINATH A., STASKO J.: VIS 25 – all the papers and citations, 2014. URL: <http://www.cc.gatech.edu/gvu/ii/citevis/VIS25/>. 2
- [YEB16] YALÇIN M. A., ELMQVIST N., BEDERSON B. B.: Keshif: Out-of-the-box visual and interactive data exploration environment. In *IEEE VIS Workshop on Visualization in Practice* (2016). URL: <http://adilyalcin.me/academic/KeshifInPractice.pdf>. 3