

# Pilaster: A Collection of Citation Metadata Extracted From Publications on Visualization for the Digital Humanities

Alejandro Benito-Santos and Roberto Therón

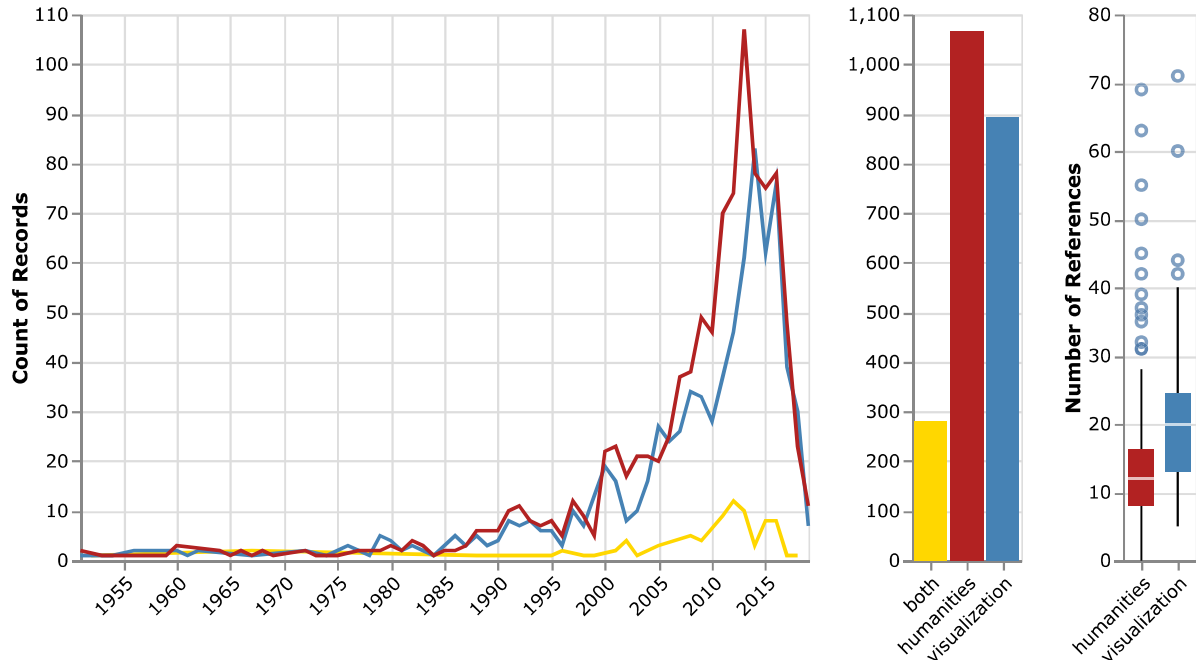


Fig. 1: Three visualizations depicting different features of the citations dataset. From left to right, the first chart shows a portion (1950–2019) of the temporal distribution of resources in the citations dataset originating in the VIS community (blue), in the digital humanities community (red), or in both (yellow). The second chart shows total record counts in the citations dataset for each of these three categories. Finally, the whisker plots on the right display the distribution of reference list lengths by publication venue for records in the seed dataset.

**Abstract**— In this paper, we present Pilaster, a collection of citation metadata extracted from publications in visualization for the digital humanities. The collection is generated from a seed set of relevant publications from which we extracted cited works, including journal and conference papers, books, theses, or blog posts, among other resources. The main aim of this work revolves around three main points: first, the collection may serve as an *entry point* to the discipline for digital humanists and visualization scholars without previous experience in the field. Second, Pilaster can be regarded as a *meeting point* for more established visualization or humanities scholars seeking to collaborate in the development of novel research ideas and related visualization design studies in the context of the humanities. Third, and given the large amount of visualization design spaces that were captured, we believe the dataset has the potential to become the *starting point* for future studies aimed at understanding the particularities of problem-driven visualization research in this and other contexts.

**Index Terms**—collaboration, dataset, digital humanities, visualization, citation analysis, scientometrics

## 1 INTRODUCTION

The collaboration between computer scientists and humanities scholars presents a highly interesting field of experimentation that has produced

- Alejandro Benito-Santos is with VisUSAL, University of Salamanca, Spain. E-mail: abenito@usal.es.
- Roberto Therón is with VisUSAL, University of Salamanca, Spain. E-mail: theron@usal.es.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxx

important learning outcomes in the past and continues to do so until today. In general, and as it has occurred in other disciplines of science, applying computational methods to humanities research workflows has helped accelerate knowledge discovery and enhance the overall quality of results in humanistic research. A significant part of these combined efforts has typically focused on the application of data visualization techniques aimed at leveraging the interaction between humanities scholars and the said computational methods, producing interesting results in different conventional areas of humanistic research, such as discourse [5], literary [8], or poetry analysis [1, 11, 12] or the browsing and sensemaking of cultural collections [20, 21].

However, the building and organization of interdisciplinary teams of experts that can produce valuable research outcomes in both the visualization and humanities domains seldom are problem-free [10, 18]. Thus, this calls for special considerations to be taken into account by

all the involved parties [14]. This is particularly the case for visualization researchers new to the field whose previous experience may lie in other areas of visualization practice, and who may rapidly become overwhelmed by the complexities of the collaboration. Analogously, humanities scholars without previous or little experience in participating in visualization design studies may also encounter problems when trying to specify requirements and tasks due to their lack of visualization literacy [15].

The results presented in this paper constitute an extension of our recent work in the field [3,4] that aims at supporting the immersion process [7] of interdisciplinary researchers in visualization design studies within a digital humanities context, among other goals that are described throughout the paper. To this end, we employ a metadata collection of works on visualization for the digital humanities that we started building in 2019 and that we have kept curating and refining since then. The resulting dataset comprises almost 2,000 resources related to the practice of visualization in the context of digital humanities derived from an extensive analysis of the citations in a core set of 119 papers published at three different venues identified at the beginning of the study. In the following sections, we discuss the rationale we followed to build the dataset, and some of the problems we found in the process and which we could not fit into our previous contribution [4] due to space limitations. Later, we present a description of the data fields and provide several descriptive statistics derived from the data that also offer new insight into the collection. Finally, we exemplify potential applications of *Pilaster* with two simple use cases that others may find useful for carrying their own studies on the dataset. The first use case aims to capture our latest work on normalizing publication aggregation names, an effort that yielded new interesting insights into the commonalities and differences in venues commonly cited by DH and VIS researchers. In a second use case, we shed new light on how collaborations in the field are articulated, which suggests a lack of overlap between the two communities.

## 2 SURVEYING VIS4DH

"Let's be honest—there is no definition of digital humanities, if by definition we mean a consistent set of theoretical concerns and research methods that might be aligned with a given discipline [...] How else to characterize the meaning of an expression that has nearly as many definitions as affiliates? It is a social category, not an ontological one."

R.C. Alvarado in *The Digital Humanities Situation* (2011) [2]

As mentioned in the previous section, the dataset arises from citations found in a core set of publications on visualization for the digital humanities. According to established literature review methodologies [13], the process of literature review starts by defining the scope of the study, ("*visualization for the digital humanities*" in our case). In the next step, the scope is condensed in a series of textual queries that are launched against online literature databases to obtain relevant publications. Then, these publications are analyzed, summarized, and discussed according to the classification dimensions and other traits derived by the authors of the survey [13]. Finally, the results of the review are wrapped up and prepared for dissemination to the scientific community. Whereas the process is seemingly straightforward, and we knew of similar methodologies that had been successfully applied to conduct surveys on specific sub-fields of the DH visualization practice [20], it presents several issues that rendered it unfit for our purpose of capturing the different DH areas in which the visualization practice *mostly* occurs. Besides, much of the work in digital humanities is presented exclusively at annual conferences (although some notable journals exist) whose proceedings are not indexed in the main online scientific databases. If, as it was our initial intention, our work should be aimed at interdisciplinary visualization practitioners, completely excluding all these works from an initial analysis seemed clearly counterproductive. Still, and beyond these considerations, we had to provide a sensible definition of the digital humanities to commence the survey. Here, we were facing a recurrent problem of the digital humanities that has been at the center of many academic debates: we resorted to the literature looking for a working definition of digital humanities that we

could put to use, but we could not find any. *How were we supposed to survey a topic that cannot be defined?* [6]

As some authors like Alvarado have pointed out, the answer for the question of what the digital humanities are cannot rely on conventional conceptions of what a discipline should be [2]. Rather, he claims, it is more useful to see the digital humanities as a *social category* that relates a collective of researchers who are involved in different, probably distant disciplines, and who call themselves "digital humanists." This statement was the cornerstone on which the methodology we adopted to generate the collection was built, and allowed us to move on to the data collection stage without the need to provide a definition of the digital humanities that would have stood on very shaky epistemological grounds, let alone a query that translated this definition into something that could be understood by a search engine. In such circumstances, we decided to adopt an utilitarian stance that focused instead on identifying the group of scholars who call themselves digital humanists and practice visualization. Taking this reasoning further forward, it seemed obvious that this group must be composed of visualization practitioners interested in digital humanities, and also of digital humanists who have shown an interest for visualization. As we discuss in the next section, we looked for specific academic collectives whose members matched any of these two conditions.

## 3 DATA COLLECTION

In this section, we detail how we built a seed dataset of publications from which the citations were extracted at a later stage. The methodology that we followed to build the dataset is inspired by other recent works in visualization research [9, 13] that were adapted to cope with the diffuse character of digital humanities, as we explain in Sect. 2.

### 3.1 Sampling VIS authors

As explained before, the construction of the seed dataset involved the sampling of publications in both ends of the humanities-visualization collaboration. To find the components of the first group, we considered participants in the last editions of the VIS4DH workshop which is, to the best of our knowledge, the only space devoted to the task of "bringing together researchers and practitioners from the fields of visualization and the humanities to discuss new research directions at the intersection of visualization and (digital) humanities research<sup>1</sup>." Although we knew of more research papers published at visualization conferences that could probably have been included in the seed dataset, we decided not to do so due to the aforementioned impossibility of establishing a well-defined boundary between what qualifies as digital humanities and what not. At any rate, we assumed relevant papers would eventually appear during the analysis of the citations and therefore we preferred to keep the seed dataset as well scoped as possible. The inspection of the proceedings of the four first editions (2016-2019) of the workshop left a balance of 47 papers and 136 authors making up the first sample to be part of the seed dataset.

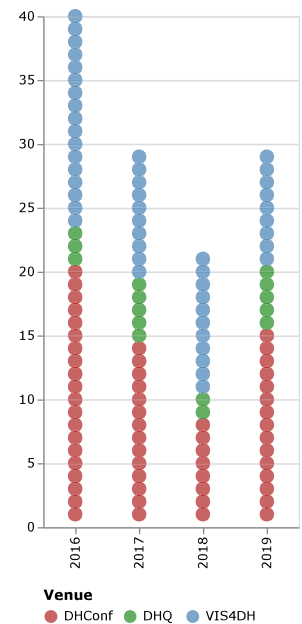


Fig. 2: Dot plot showing the distribution by year and publication venue of papers in the seed dataset. A total of 119 papers were analyzed in a first stage.

<sup>1</sup><https://vis4dh.dbvis.de/>

### 3.2 Sampling DH authors

To obtain representative publications in the humanities side, we decided to inspect the proceedings of the last 4 editions (2016-2019) of the joint annual conference of the Alliance of Digital Humanities Organizations (ADHO) and its peer journal Digital Humanities Quarterly (DHQ). However, and unlike the previous case, we could not find a similar event to the VIS4DH in the DH Conference, and rather visualization practice seems to be spread across different areas such as geohumanities, linked open data, or audiovisuals<sup>2</sup>. Given that we wanted to capture all works that employed visualization techniques regardless of their area of application, we opted for capturing long presentations and papers in the two venues that were related to visualization as tagged by their own authors. Concretely, we captured publications whose title, user-authored keywords or list topics (topics are chosen by the authors from a list of keywords compiled by ADHO) matched the regular expression “[Vv]isua\*”. The search yielded a total of 72 publications (57 long presentations from the conference proceedings and 15 long papers from the journal) which constituted the “humanities” part of the seed dataset. The final composition of the seed dataset is shown in Fig. 2.

## 4 DATA PROCESSING

Publications in the first group were downloaded in PDF format from the workshop’s homepage and their respective reference lists extracted with the `pdftotext`<sup>3</sup> library and stored for later processing. Reference lists of the second group were obtained by parsing the TEI-XML files in which the documents were encoded. The TEI files of publications in the DH Conference proceedings and the DHQ journal were obtained from the ADHO’s GitHub repository<sup>4</sup> and from the journal’s website, respectively. The TEI files of the 2019 edition of the DH Conference had to be directly scraped from the conference website as they were missing from the repository. The bibliography sections of each paper in the seed dataset were analyzed with the Neural-ParsCit suite [16], which automatically extracts diverse metadata from text lines in a paper’s reference list. The metadata includes but it is not limited to the title, publication year and venue, authors list, DOI and URL. For each of the extracted works, we completed their metadata with information obtained from the Elsevier API<sup>5</sup> by matching their name with existing records in the database. Finally, author names and publication venues were normalized by following a semi-supervised iterative procedure that consisted in visually inspecting pairs displaying short edit distances. Whenever the names were found to refer to the same entity (author or venue), they were unified under their most common form. This process was repeated until no similar pairs were left. At the end of the extraction process, we obtained 2238 references of works that were cited from the seed dataset. They were resolved to 1934 different works of which 23 were publications originally included in the seed dataset.

## 5 DATASET DESCRIPTION

In this section, we describe the data fields that compose the entries and provide general descriptive statistics of the values they take. Below, we list data attributes that are common to items found in the seed or citations datasets:

- **key:** An automatically generated random key that identifies a given resource.
- **title:** The resource title obtained
- **authors:** The item’s list of authors separated by semicolons. The complete list of authors comprises 3499 names of which 185 can also be found in the seed dataset.
- **aggregation:** The normalized name of the aggregation in which the item can be found (e.g., a conference names or journal/book titles). We identified 1148 different aggregation types holding 1783 items in the citations dataset.

<sup>2</sup><https://adho.org/special-interest-groups-sigs>

<sup>3</sup><https://github.com/jalan/pdftotext>

<sup>4</sup><https://github.com/ADHO>

<sup>5</sup><https://dev.elsevier.com/>

- **year:** The year in which the item was created. It was obtained by parsing the reference or from the Elsevier API.
- **source\_theme:** Denotes the provenance of the record. For items in the seed dataset, this field takes two values (“visualization”, “humanities”) depending on the type of the sample that included them, as described in Sect. 3. The value is inherited by items in the citations dataset to annotate their provenance. Items cited from both parts of the seed dataset have this value set to “both”.

Additionally, items in the **seed** dataset contain the following three extra fields:

- **publication\_short\_title:** An abbreviated form of publication\_title.
- **author\_keywords:** Keywords list given by the items’ authors.
- **n\_references:** Length of the reference list that can be found at the end of the paper.

Finally, data attributes exclusive to items in the **citations** dataset are listed below:

- **cited\_by:** A list of foreign keys pointing to papers in the seed dataset that cite the item.
- **cited\_by\_venue:** Venue (VIS4DH, DH Conference, DHQ) of the paper(s) citing the item.
- **cited\_by\_count:** Number of papers in the seed dataset that cite a given item excluding self-references. We considered a citation to be a self-reference when the set intersection between the authors of the citing work and the authors of the cited work was not the empty set.
- **type:** In cases where the publication could not be matched again an Elsevier record, we derived its type (e.g., conference paper, journal article, book) from other publications in the same venue that could be found. In total we identified 20 different cited work types (Fig. 3).
- **aggregation\_type:** The type of the aggregation, if existent, in which the item can be found (e.g., journal, conference proceeding, or book).
- **link:** Web links extracted from the original reference that were parsed by means of a regular expression.

In Fig. 1, we present some descriptive statistics that give an idea of the composition of the citations dataset according to its different dimensions. The first chart on the left shows how both communities follow similar temporal citation patterns with similar mean (2006) and median (2011) values. The next chart shows how the cited resources can be divided into three groups according to the community their citing counterparts belong to. As it can be seen in the figure, we obtained 280 resources that were referenced from VIS4DH and DHConference/DHQ papers, which in turn are among the most cited in the dataset: 82 out of the 100 most cited works belong to this category, which were cited a total of 267 times (11.93% of all citations by papers in the seed dataset). Publications in this category are highly relevant because they represent the intersection point between the visualization and DH communities and therefore, they describe a shared communication channel [18] between visualization and domain experts in DH research that we believe it is worth studying in greater depth. The seed and citations dataset were stored in a public spreadsheet<sup>6</sup> for ease of use by other researchers.

## 6 USE CASES

In this section we propose two simple use cases of the dataset that can help to illustrate the potential use cases for the dataset. The first use case employs the citations dataset to explore commonly cited venues. Besides, we show other venues that are cited exclusively by researchers in one of the two sides. The second use case provides some insights on how interdisciplinary teams are conformed and how the collaborations are organized. The figures in this and the other sections of the paper

<sup>6</sup>[https://docs.google.com/spreadsheets/d/1Z8aMhxpai510hkuSVAFW6L4QyQfPPvUnv8IjKuF2\\_Jo/](https://docs.google.com/spreadsheets/d/1Z8aMhxpai510hkuSVAFW6L4QyQfPPvUnv8IjKuF2_Jo/)

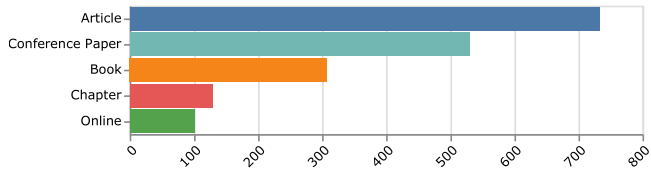


Fig. 3: Five most common resource types in the citations dataset. A majority of the cited items (1,263, 65.30%) belong to one of the two top categories, although there are also references to books (307), book chapters (128) or online resources (100), such as blog posts or datasets.

were generated in Python code <sup>7</sup> using the Vega-Lite grammar [17] and Altair [19].

### 6.1 Studying publication aggregations

In this first use case, we are interested in exploring what venues are cited most often from what kinds of sources in the seed dataset. The stacked bar chart in Fig.4.a shows aggregations above the 95th percentile by number of times cited. From this visualization, some information can be decoded: for example, the two tallest bars in the chart depict the top two most cited venues, which are *IEEE TVCG* and the *DH Conference proceedings*. Moving to the right of the chart, other venues typically associated with visualization research appear, such as the *Conference on Human Factors in Computing (CHI)*, *Computer Graphics Forum*, and the *VIS4DH* workshop, all of which are cited more or less evenly from the two categories of the seed dataset. A similar effect happens with other venues typical of DH research, among which we can find *Digital Humanities Quarterly*, *Digital Scholarship in the Humanities* and its previous title, *Literary and Linguistic Computing*. As opposed to VIS venues, there seems to be a larger imbalance between the categories of items citing DH venues, which are mostly from works in the DH seed dataset. Closer to the tail of the distribution, we can detect other special venues that are *exclusively* cited by publications originating in the DH domain, such as the *International Conference on Document Analysis and Recognition (ICDAR)* or the *Annual Meeting of the Association for Information Science and Technology (ASIS&T)*. We capture this idea in more detail in the chart of Fig.4.b, which represents venues cited exclusively by at least two publications in one of the two domains. These venues, we argue, may be indicative of current knowledge gaps in both sides of the visualization practice that could point to potential new areas for collaboration.

### 6.2 Exploring the authors graph

In this second use case, we obtain insight into the size and structure of collaborations by means of a social network analysis of co-authorship relationships found in the seed dataset. The node-link diagram of Fig. 5 depicts collaborations in both areas. By looking at the color of nodes in the chart, it can be seen that the number of authors who published papers in both categories is fairly (2.76%) low, meaning that interactions between the two communities still are scarce, a fact that may be linked to certain issues pointed by other authors in the past [10]. Attending to the topology of the graph, author communities in the VIS side appear to be larger than their counterparts in the DH side, which may be partially due to differences in the average number of authors per paper in the two groups (DH:  $2.92 \pm 1.51$  vs VIS  $4.15 \pm 2.10$ ) but probably also to other factors that may deserve further study.

## 7 LIMITATIONS AND FUTURE WORK

### 7.1 Data Collection Methodology

In Sect. 3, we described the rationale that we followed to sample publications in both sides of the collaboration. Although we made an explicit effort to obtain a set of publications that was representative of the discipline, we are aware that, due to certain characteristics of the

<sup>7</sup><https://colab.research.google.com/drive/15cNprIDXS1WMa660lo-ApinMib8vdth>

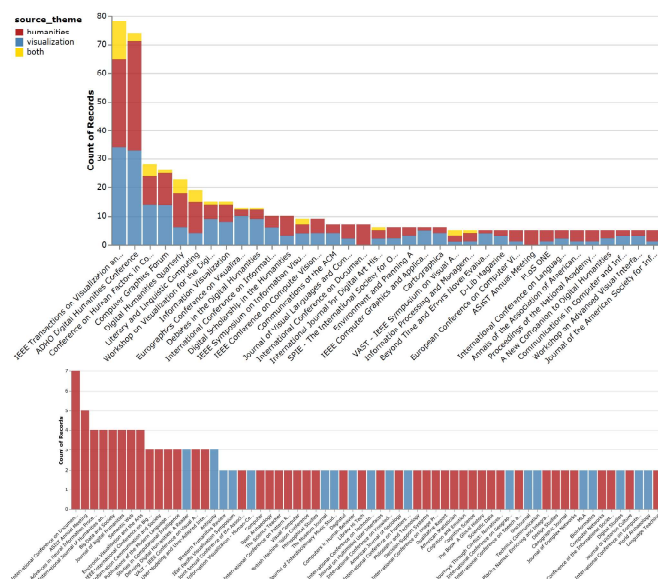


Fig. 4: (a) Top 5% most cited aggregations in the citations dataset. IEEE Transactions on Visualization and Computer Graphics and the ADHO DH Conference are the most cited. (b) Most popular aggregations that are referenced exclusively by at least two different works in the VIS or DH seed datasets.

employed methodology, we might have missed previous work that could have been part of the seed dataset. For example, this could happen with VIS authors working on DH topics who have not participated in the VIS4DH workshop. A similar effect could happen with DH practitioners who decided not to include any terms matching the regular expression “[Vv]isua\*” in their abstracts. In this respect, we expect to receive suggestions from the community of potential new sources that can be included as part of the collection in future developments to make it more complete.

### 7.2 Differences in Publication Formats

The distribution of citations according to their provenance is skewed towards the humanities side, a phenomenon that can be traced to differences between the publication formats typically used on each domain. For example, whereas long presentations at the DH Conference are submitted as abstracts of maximum 750-1000 words, submissions to the VIS4DH workshop adopt the short paper form of 4+1 pages, which usually yield around 3500-4000 words ( $\approx 3x$  longer). Although this difference in length is not translated into a similar difference in the average number of citations per paper between the two categories (right of Fig. 1), humanities papers consistently generated less citations on average than their VIS counterparts. However, they represent a thematically richer set of publications. Although we believe this fact is just representative of the reality of the field and it is not a drawback in itself, it is important to take it into account before extracting any conclusions from the dataset.

### 7.3 Head or Tails

In this paper, we tried to provide an overview of the collection by focusing on the heads of the rank-frequency distributions of, for example, resource types (Fig. 3) or publication aggregations (Fig. 4). Whereas we believe this kind of analysis serves well the objective of describing the dataset, we are aware that this practice may also have unintended side effects: for example, it could happen that these rank-frequency distributions may be interpreted as importance rankings that go beyond the purpose of providing an entry point to the dataset, a practice which we have argued against in the past [3]. By looking *only* at top-ranked items while disregarding the rest, other vital information for advancing

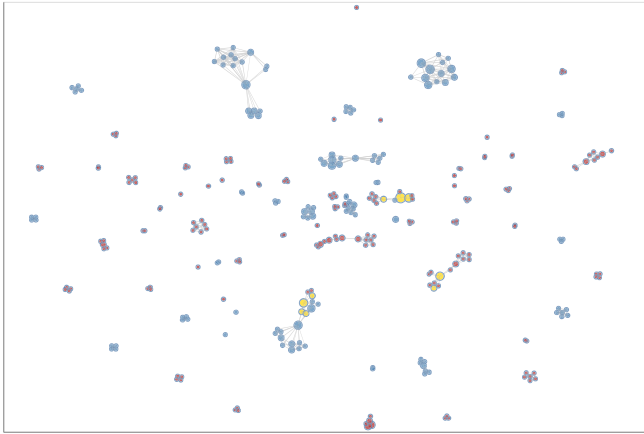


Fig. 5: A node-link diagram depicting co-authorship relationships between authors in the seed dataset. Only 9 out of 328 (2.76%, in yellow) of individuals authored publications in both the VIS and the DH datasets.

the field may be missed, a practice that also dangerously contributes toward perpetuating prestige bias (among other biases) in academia. Rather, we recommend potential users of the collection to repair on items found at the tails of the distributions, for example by performing searches on specific terms that could unveil highly-interesting but lowly-cited, underrepresented themes, works, venues, or authors.

## 8 CONCLUSION

In this paper, we presented *Pilaster*, a metadata collection of papers and related citations employed by scholars working at the intersection of visualization and digital humanities. By departing from a representative sample of publications in the field, we aimed at capturing the different perspectives of scholars at both ends of the collaboration. Furthermore, we exemplified how insight into the discipline can be obtained by means of two use cases that can be easily adapted by other researchers to cover more complex interactions and usage scenarios. In addition, the resulting spreadsheet and code used to generate the figures in this paper were put in the public domain and can be consulted online. Beyond serving as an entry point to the discipline for novel researchers to the field, the results of our work are also aimed at more established scholars who may find them useful for detecting potential future collaborations or novel research ideas, as we illustrated in Sect. 6. Although we plan to continue updating the dataset as new publications become available, we encourage other researchers to send us feedback or suggestions of other use cases that we may not have covered here.

## REFERENCES

- [1] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. Rule-based Visual Mappings – with a Case Study on Poetry Visualization. *Computer Graphics Forum*, 32(3pt4):381–390, June 2013. doi: 10.1111/cgf.12125
- [2] R. C. Alvarado. The digital humanities situation. *Debates in the digital humanities*, pp. 50–55, 2012.
- [3] A. Benito-Santos and R. Therón Sánchez. Cross-domain Visual Exploration of Academic Corpora via the Latent Meaning of User-Authored Keywords. *IEEE Access*, 7:98144–98160, 2019. doi: 10.1109/ACCESS.2019.2929754
- [4] A. Benito-Santos and R. Therón Sánchez. A Data-Driven Introduction to Authors, Readings and Techniques in Visualization for the Digital Humanities. *IEEE Computer Graphics and Applications*, pp. 1–1, 2020. doi: 10.1109/MCG.2020.2973945
- [5] M. El-Assady, V. Gold, A. Hautli-Janisz, W. Jentner, M. Butt, K. Holzinger, and D. A. Keim. VisArgue : A Visual Text Analytics Framework for the Study of Deliberative Communication. In *PolText 2016 - The International Conference on the Advances in Computational Analysis of Political Text*, pp. 31–36, 2016.

- [6] M. K. Gold. Day of dh: Defining the digital humanities. *Debates in the Digital Humanities*, pp. 69–71, 2012.
- [7] K. W. Hall, A. J. Bradley, U. Hinrichs, S. Huron, J. Wood, C. Collins, and S. Carpendale. Design by Immersion: A Transdisciplinary Approach to Problem-Driven Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):109–118, Jan. 2020. doi: 10.1109/TVCG.2019.2934790
- [8] U. Hinrichs, S. Forlini, and B. Moynihan. Speculative Practices: Utilizing InfoVis to Explore Untapped Literary Collections. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):429–438, Jan. 2016. doi: 10.1109/TVCG.2015.2467452
- [9] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. D. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. Vispubdata.org: A Metadata Collection About IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2199–2206, Sept. 2017. doi: 10.1109/TVCG.2016.2615308
- [10] S. Jänicke. Valuable Research for Visualization and Digital Humanities: A Balancing Act. In *Proc. 1st Workshop on Visualization for the Digital Humanities (VIS4DH)*, 2016.
- [11] S. Jänicke and D. J. Wrisley. On Alignment of Medieval Poetry. In *Digital Humanities 2018 Book of Abstracts*, 2018.
- [12] N. McCurdy, J. Lein, K. Coles, and M. Meyer. Poemage: Visualizing the Sonic Topology of a Poem. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):439–448, Jan. 2016. doi: 10.1109/TVCG.2015.2467811
- [13] L. McNabb and R. S. Laramée. How to Write a Visualization Survey Paper: A Starting Point. In M. Tarini and E. Galin, eds., *Eurographics 2019 - Education Papers*. The Eurographics Association, 2019. doi: 10.2312/eged.20191026
- [14] M. Miller, H. Schäfer, M. Kraus, M. Leman, D. A. Keim, and M. El-Assady. Framing Visual Musicology through Methodology Transfer. *Proceedings of the Workshop on Visualization for the Digital Humanities (VIS4DH) at IEEE VIS 2019*, Oct. 2019.
- [15] J. Pereda, P. A. Murrieta Flores, D. R. Panagiotis, and J. C. Roberts. Tangible User Interfaces as a Pathway for Information Visualisation for Low Digital Literacy in the Digital Humanities. In *Proc. 2nd Workshop on Visualization for the Digital Humanities (VIS4DH)*, 2017.
- [16] A. Prasad, M. Kaur, and M.-Y. Kan. Neural parsit: A deep learning based reference string parser. *International Journal on Digital Libraries*, 19:323–337, 2018.
- [17] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, Jan. 2017. doi: 10.1109/TVCG.2016.2599030
- [18] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair. Bridging the gap of domain and visualization experts with a Liaison. In *Accepted at the Eurographics Conference on Visualization (EuroVis)*, vol. 2015. The Eurographics Association, Cagliari, Italy, 2015.
- [19] J. VanderPlas, B. Granger, J. Heer, D. Moritz, K. Wongsuphasawat, E. Lees, I. Timofeev, B. Welsh, and S. Sievert. Altair: Interactive Statistical Visualizations for Python. *The Journal of Open Source Software*, 3:1057, Dec. 2018. doi: 10.21105/joss.01057
- [20] F. Windhager, P. Federico, G. Schreder, K. Glinka, M. Dörk, S. Miksch, and E. Mayr. Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018. doi: 10.1109/TVCG.2018.2830759
- [21] F. Windhager, S. Salisu, and E. Mayr. Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections. *Informatics*, 6(3):29, Sept. 2019. doi: 10.3390/informatics6030029