

Visualizing DHQ (Digital Humanities Quarterly) Bibliography

Mapping Cultures in the Big Tent: Multidisciplinary Networks in the Digital Humanities Quarterly

Authorship

Dulce Maria de la Cruz
Freelance Data Analyst

Dulce.Maria.delaCruz@gmail.com

Kristin Lewis
AAAS/NASA

kristin.l.m.lewis@gmail.com

Teh-Hen Yu
IT Professional

tehhenyu@hotmail.com

Jake Kaupp
Engineering Education Researcher
Queen's University
Kingston, ON, Canada
jkaupp@gmail.com

Max Kemman
PhD Candidate
University of Luxembourg
Luxembourg
maxkemman@gmail.com

Description of visualization goal/need, hand-sketch of the envisioned visualization, and discussion why this project is important

Goal/Need

The Digital Humanities Quarterly (DHQ) journal covers all aspects of digital media in the humanities, representing a meeting point between digital humanities research and the wider humanities community [1]. It is the publication from ACH (Association for Computers and Humanities), which is part of ADHO (the Alliance of Digital Humanities Organization) -- a global alliance with constituent members in EU (EADH), US (ACH), Canada (CSDH/SCHN), Japan (JADH), Australia (aaDH), and an international network (centerNet) with 196 Digital Humanity Centers globally. Articles published in DHQ involve authors of multiple countries, institutions and disciplines who work on several subjects and areas related to digital media research.

Under a recent grant from NEH (National Endowment for Humanities), DHQ has developed a centralized bibliography which supports the bibliographic referencing for the journal. The client is looking for visualizations that show:

1. how citations reflect differences in academic culture at the institutional and geographic level
2. the changes to that culture over time.

3. correlations between article topics (reflected in keywords) and citation patterns.

The identification of those subjects and areas and of their major contributors would be very important for any researcher involved or interested in digital media research. However, due to the collaborative, multidisciplinary nature of the digital media research, such identification becomes extremely difficult, if not impossible, to accomplish by merely analyzing the DHQ bibliographic database [2]. In such a case, visualization is the preferred approach.

Hand-sketch

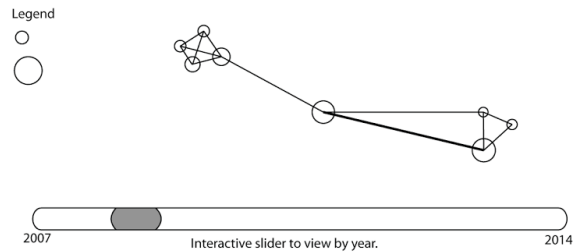
Taking into account the available data for each DHQ article, we have planned a visualization that includes several components which may reveal clusters representing different cultures within this highly interdisciplinary research. First, we will create a bibliographic coupling of authors. The usual approach of bibliographic coupling is to cluster articles [3], we however deviate from this for two reasons: 1) we are interested in cultures, which is best represented by people rather than articles, and 2) due to the interdisciplinary nature of DH a single article with multiple authors can represent multiple cultures (e.g. a collaboration between a historian and a computer scientist). Second, we will create a co-citation network of cited articles [4].

For each cluster, we will include a word cloud in order to illuminate correlations between articles topics (reflected in keywords) and citations patterns. An interactive visualization with a slider by year (or an automated animated visualization) will help reveal how these clusters change or grow over time. Potential visualizations are shown in the diagrams below:

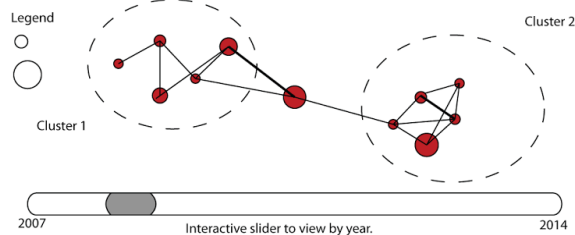
Mapping Cultures in the Big Tent: Multidisciplinary Networks in the Digital Humanities Quarterly

Description/Key Findings

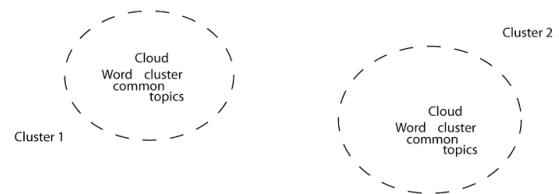
Bibliographic Coupling of Authors



Paper Co-citation Networks



Keywords per Cluster



The suggested visualization aims to allow DHQ's readers to gain insight into the citation networks that operate within the journal, showing them the major contributors and the subjects and areas involved in digital media research.

Dataset

Two tables were extracted from the Client dataset:

1. dhq_articles (178 records with unique columns: article_id, author, year, title, journal/conference/collection, abstract, reference IDs, isDHQ)
2. works_cited_in_dhq (3823 records; same columns as above table)

We initially encountered problems in the data, which required some pre-processing, pruning and reconciliation of the datasets. For instance:

- Incomplete, inconsistent and missing records in the dataset.
- There is no author institution affiliation available for both tables. That makes it quite difficult if we are to group the authors by institutions or GEOs.

- Keywords are only provided for a subset of the articles in the xml information.

The first issue will be easily remedied by contacting the client to verify the completeness of the provided data, and seeing if they can address any gaps; or by developing a consistent framework to exclude inconsistent or incomplete records. We will also cross reference the records with other information provided in the dataset, to ensure the completeness and accuracy of the dataset.

The second issue requires additional work, as the provided information is not sufficient to identify the department or disciplinary affiliation of the author. In addition, there is no geographic information or institutional affiliation in the dataset. In our group's initial efforts to address the second issue we were able to scrape the DHQ web site (<http://www.digitalhumanities.org/dhq/index/author.html>) to extract information about the affiliation of the first author for each DHQ article; this affiliation was then extended to provide geospatial information for the primary author. Affiliation data (at an institutional level) is also available in provided XML files, which requires additional processing. Continuing efforts will focus on identifying the departmental and disciplinary affiliation of the authors by search engines such as Google, Google Scholar, Web of Science and the scholarly database provided by the Cyberinfrastructure for Network Science center at Indiana University.

We also wish to contact the client, as the initial description of the project includes mention to additional items not included in the provided dataset, such as topic keywords for each paper. However, we can conduct topical analysis using provided abstract or title fields if keywords are not available.

Importance

Digital Humanities (DH) is a field of research difficult to define due to its heterogeneity¹. With its inclusionary ambitions, DH is regularly referred to as a 'big tent' [5] encompassing scholars from a wide variety of disciplines such as history, literature, linguistics, but also disciplines such as human-computer interaction and computer science. This collaborative, multidisciplinary approach to digital media makes DH an interesting field, but also difficult to grasp. A question is to what extent the big tent of DH represents a single epistemic culture, or actually a variety of cultures [5, 6].

DHQ is arguably one of the largest journals specifically aimed at DH research. As such, it has attracted publications from across the big tent. To gain an understanding of the diversity of culture(s) in the DH, we are interested in how unique disciplinary cultures are represented in the DHQ. Considering cultures are

¹ See e.g. <http://whatisdigitalhumanities.com/> for a wide variety of definitions from different scholars

self-referential systems, we might expect that scholars from a certain culture are more likely to cite scholars from their own discipline rather than from others [6]. As such, we expect citation behaviour to reflect disciplinary cultural norms. Therefore, visualizing and analysing the bibliographic data of DHQ not only gives insights into the specific bibliographies from DHQ, it might give insight into the way the different epistemic cultures in the DH big tent interact with one another, and how this interaction and collaboration impacts the networks over time.

Discussion of related work

Visualization of Citation Networks

Ever since Garfield's work on citation analysis [7], citations have been of interest to understand how scientists and scholars build upon one another's work. Generally, citations are a useful metric for understanding academia because they constitute the simplest relation between two publications [8, 9], representing that the authors have read and been influenced by another paper [10]. Citations are thus a useful way to understand the relations of a publication to previous work and its impact on subsequent research [11]. As such, understanding how knowledge diffuses through academic cultures can be approached by analysing citations. This was done manually in the past, but has become much more feasible and scalable thanks to computational analysis [12]. By showing citations as a graph, documents can be understood as part of a citation environment [9]. Clusters of citation environments may represent cultures of knowledge present in the journal. Therefore, we will cluster the citation data as described above in "Hand-sketch".

Many of these visualizations are build by scientists from Bibliometrics or Science and Technology Studies. However, there is an increasing interest in visualization also from DH scholars [13]. A DH project that aimed at humanities scholars that visualized bibliographies was RoSE (Research-oriented Social Environment) that aimed to "storyboard" intellectual movement [14]. Likewise, our visualization will be aimed at humanities scholars.

Academic Cultures

Culture in higher education is both broadly used and difficult to define. It can pertain to many different social structures within and across an institution, providing many different units of analysis [15]. One of the most commonly studied units is the the departmental or disciplinary sub-culture, which can be represented equally

across institution, country and geographic boundaries [16]. There is no universal definition and depiction of culture but it is typically comprised of common beliefs and perceptions that coalesce around similar values of norms of groups of individuals of common background, training and experience [17]. These values and norms comprising disciplinary cultures can be separated into, first; the epistemic (“ways of knowing and organizing knowledge”) cognitive aspect consists of the general topical area of expertise and established research methods and resources [12]. Second, the socio-organizational, hierarchical aspect which is predominantly defined by institutional structures and divisions, such as departments or faculties that comprise an institution [12].

With respect to our data, we can categorize or classify the topic of each paper according to a specific institution through textual analysis of abstracts or titles. Consistent with this approach, we can also identify the departmental affiliation of author through deeper research and web-scraping. This permits the classification of topic, author, institution and departmental affiliation, providing an approach to defining the cultural identity of dataset records. However, this approach is not without limitations and drawbacks with the most prominent being that departments can contain multiple disciplinary cultures [18]. This highlights the importance of focusing our classification along epistemic lines, which embody the definition set forth by Knorr Cetina:

“those sets of practices, arrangements and mechanisms bound together by necessity, affinity and historical coincidence which, in a given area of professional expertise, make up how we know what we know. Epistemic cultures are cultures of creating and warranting knowledge” [19, p363].

By using this more malleable concept, we are able to use the publication history of an author alongside the departmental and institutional categories to produce a more holistic classification scheme which considers both the institutional disciplinary organization (the department) along with the knowledge-based classification of the discipline. We believe that this approach will permit the classification of different cultures within a journal as multidisciplinary and collaborative as the Digital Humanities Quarterly, even if different disciplinary cultures exist within a single department, according to citation behaviour.

References

- [1] Digital Humanities Quarterly (n.d.). *About DHQ*. Retrieved from <http://www.digitalhumanities.org/dhq/about/about.html>
- [2] Borner, K., Chen, C. M., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review Of Information Science And Technology*, 37, 179–255. doi:10.1002/aris.1440370106
- [3] Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American documentation*, 14(1), 10-25.

- [4] Weingart, S. (2013) Networks Demystified 4: Co-Citation Analysis. *the scottbot irregular*.
<http://www.scottbot.net/HIAL/?p=38272>
- [5] Svensson, Patrik. (2012) Beyond the big tent. *Debates in the Digital Humanities*, 36-49.
- [6] Knorr Cetina, K. (2007). Culture in Global Knowledge Societies: Knowledge Cultures and Epistemic Cultures. *The Blackwell Companion to the Sociology of Culture*, 32(4), 361–375.
doi:10.1002/9780470996744.ch5
- [7] Garfield, E. (1955). Citation indexes for science. A new dimension in documentation through association of ideas. *Science*, 122(3159), 108–111.
- [8] Brüggemann-Klein, A., Klein, R., & Landgraf, B. (1999). BibRelEx: exploring bibliographic databases by visualization of annotated content-based relations. *2000 IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics*, 5(11), 19–24.
doi:10.1109/IV.2000.859731
- [9] Lu, W., Janssen, J., Milios, E., Japkowicz, N., & Zhang, Y. (2006). Node similarity in the citation graph. *Knowledge and Information Systems*, 11(1), 105–129. doi:10.1007/s10115-006-0023-9
- [10] Elmqvist, N., & Tsigas, P. (2007). CiteWiz: a tool for the visualization of scientific citation networks. *Information Visualization*, 6(3), 215–232. doi:10.1057/palgrave.ivs.9500156
- [11] Schäfer, U., & Kasterka, U. (2010). Scientific authoring support: a tool to navigate in typed citation graphs. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids* (pp. 7–14). Association for Computational Linguistics. Retrieved from <http://portal.acm.org/citation.cfm?id=1860659>
- [12] Becher, T., & Parry, S. (2005). The Endurance of the Disciplines. In I. Bleiklie & M. Henkel (Eds.), *Governing Knowledge* (Vol. 9, pp. 133–144). Springer. doi:10.1007/1-4020-3504-7
- [13] Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1), 1–21. Retrieved from <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
- [14] Liu, A., et al. (2012) Friending the Humanities Knowledge Base: Exploring Bibliography as Social Network in RoSE. *White paper for the NEH Office of Digital Humanities*. Retrieved from <https://rosedocumentation.files.wordpress.com/2012/07/rose-white-paper-as-submitted-to-neh.pdf>
- [15] Tierney, W. (1988). Organizational Culture in Higher Education: Defining the Essentials. *The Journal of Higher Education*, 59(1), 2–21. doi:10.2307/1981868
- [16] Tierney, G., & William, J. E. (2011). Culture Organizational in Higher Education. *The Journal of Higher Education*, 59(1), 2–21. Retrieved from <http://www.jstor.org/stable/1981868> .
- [17] Martin, Joanne. *Cultures in Organizations: Three Perspectives*. Oxford University Press, 1992.
- [18] Tuunainen, J. (2005). When disciplinary worlds collide: The organizational ecology of disciplines in university department. *Symbolic Interaction*, 28(2), 205–228. doi:10.1525/si.2005.28.2.205
- [19] Knorr Cetina, K. (2007). Culture in Global Knowledge Societies: Knowledge Cultures and Epistemic Cultures. *The Blackwell Companion to the Sociology of Culture*, 32(4), 361–375.
doi:10.1002/9780470996744.ch5