
The Intellectual Structure of Digital Humanities: An Author Co-Citation Analysis

Jin Gao

jin.gao.13@ucl.ac.uk

UCL Centre for Digital Humanities, United Kingdom

Oliver Duke-Williams

o.duke-williams@ucl.ac.uk

UCL Centre for Digital Humanities, United Kingdom

Simon Mahony

s.mahony@ucl.ac.uk

UCL Centre for Digital Humanities, United Kingdom

Melanie Ramdarshan Bold

m.bold@ucl.ac.uk

University College London, United Kingdom

Julianne Nyhan

j.nyhan@ucl.ac.uk

UCL Centre for Digital Humanities, United Kingdom

Introduction

With many ongoing debates (Gold, 2012) and “unwritten” histories (Nyhan and Flinn, 2016), the research practice of the Digital Humanities (DH) has been around for 70 years. Many works have been trying to draw general conclusions of the disciplinary structure (McCarty, 2003; Gold, 2012; Terras et al., 2013; Schreibman et al., 2016; Nyhan and Flinn, 2016), and have pointed to the potential usefulness to analyse the discipline from statistical aspects. The usefulness focuses on describing the intellectual structure, scholarly interactions and disciplinary development of DH. Some studies have dedicated their attention to these matters (Grandjean, 2016; Nyhan and Duke-Williams, 2014; Quan-Haase et al., 2015; Wang and Inaba, 2009), or have focussed on one of these topics (Sugimoto et al., 2013), but few of them have engaged either with the bibliometric network method, or with the latest large-scale scholarly datasets to study the DH community as a whole.

Therefore, to fill this gap, based on a provisional dataset that has been compiled from core DH journals, this study performs an exclusive all-author co-citation analysis (ACA) with the 200 most cited scholars by fractional citation count to map and

demonstrate the intellectual structure and to identify the most influential scholar groups and topics within DH.

To the best of our knowledge, this study is the first to apply bibliometric methods to visualise DH knowledge structure and the scholar clusters. This research output will make a valuable contribution to the current discussions and debates about DH knowledge structure and wider scholarly networks.

Methodology

With ACA as the main methodology, the research contains four steps, and each with a different methodology: building a DH citation index according to the publications of these journals; selecting authors as the core objects for citation analysis; assigning scholars to different distance-based clusters by calculating the author co-citation matrix to similarity matrix (Waltman and van Eck, 2013); finally, visualising the DH citation network which aims to show the scholar clusters, and the knowledge structure and diffusion of DH.

The three DH core journals that our dataset has been constructed from are: “*Computers and the Humanities*” (*CHum*), “*Digital Humanities Quarterly*” (*DHQ*), “*Literary and Linguistic Computing*” (*LLC*) (now “*Digital Scholarship in the Humanities*”) (*DSH*). The bibliographies as well as the metadata of all their publications (including the reviews and editorials etc.) published until June 2016 have been collected. It should be noted that none of these journals spanned the whole period selected (1966-2016): *CHum*, the first DH journal started in 1996, and ceased publication in 2004; *LLC/DSH* began in 1986; *DHQ* began in 2007. Figure 1 shows the total publications each year from 1966 until June 2016 for these journals.

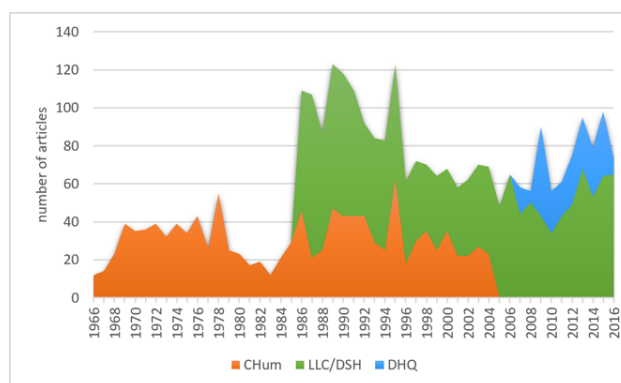


Figure 1. In total, 3,068 journal articles: *CHum* (1,195 articles with 26,033 citations), *LLC/DSH* (1,633 articles with 28,501 citations), and *DHQ* (240 articles with 4,289 citations)

Author co-citation analysis (ACA) can reveal the intellectual structure of a field from its academic publications by calculating the frequencies with which two authors are cited together. That is to say, if an article cites at least one article of author A, and at least one of author B that is different from the one of A, the co-citation count increases by 1. The more co-citations two authors receive, the more likely their publications and researches are related (Bellardo, 1980). Therefore, the clusters of related authors indicate the networks of research topics, or influential focuses within a discipline.

The initial findings with the top cited 200 authors displayed on the maps (see the provisional maps in Figure 2 and Figure 3) have provisionally revealed five sub-fields within DH.

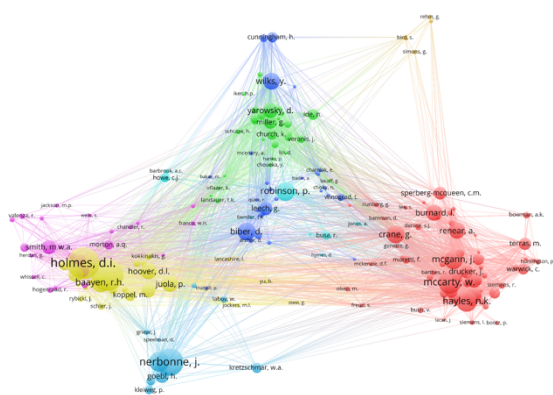


Figure 2. The provisional ACA network map in DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2016, created using VOSviewer

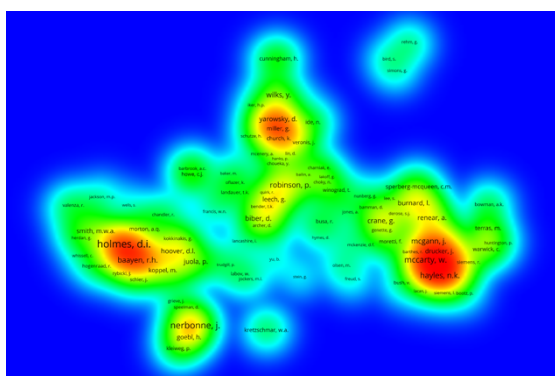


Figure 3. The provisional ACA density map in DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2016, created using VOSviewer

Both of the maps (Figure 2 and Figure 3) are distance-based. Each node on the map represents an author, and the distance between two authors is their relations (the closer the distance, the stronger the connection). Authors are distributed quite

unevenly, and this makes it easy to identify clusters of related nodes. The size of the node represents the citation count this author received, and the higher the citation count is, the bigger the node. On the density map, the density value depends on the size, number and distance of the nodes around it, so the higher the density value, the colour is more red than blue.

Both maps have revealed the general structure of the scholarly communication between DH scholars via publications. Horizontally across the centre of the map, there is a loosely connected circle of five DH scholar clusters: centre (focused on “Leech, G”), top (focused on “Miller, G”), bottom (focused on “Nerbonne, J”), left (focused on “Holmes, D.I”), and right (focused on “McCarty, W”). The clusters distribution on the density map reveals that there is a clear separation between top, centre, right clusters to left and bottom clusters. Especially the right cluster (focused on “McCarty, W”) and the left cluster (focused on “Holmes, D.I”) turn out to be denser than other clusters. This shows that these two clusters are more significant and have more citation influence. According to the provisional analysis, these five clusters appear to be associated with five different DH research topics: English study at the centre; general historical literacy and information science on the right; language modelling and natural language processing at the top; statistics and text analysis on the left; computational linguistics particularly on Dutch and German speaking at the bottom. These five clusters, however, are also grouped into two different bigger groups. The English study, language modelling, and general historical literacy seem to be in one group which is more related, while the statistics and Dutch-German linguistics are also very closely related to each other.

Limitations and Future study

This research is part of the first author’s ongoing PhD study, funded by UCL ORS scholarship and based at the UCL Centre for Digital Humanities. The doctoral research maps DH intellectual, social and environmental structures using the Invisible College model (Zuccala, 2006).

There are some limitations that need to be noted, such as the citation lag time. In order to build up a citation record for co-citation, it takes around five to eight years (Hopcroft et al., 2004). This could explain that certain recognisable authors might not appear on the maps yet. Also, because the co-citation method studies the knowledge base as its subject, the map emphasises more on authors published some time ago, which might not include the “new comers”.

In the future work, the ACA study will be extended to include more citation data. The ACA

study will be divided into discreet periods to construct maps of different DH development stages. Given that different journals have different topical foci, the research will also analyse individual journal to discover its attribute.

Bibliography

- Bellardo, T.** (1980). "The use of co-citations to study science." *Library Research*, 2(3): 231–237.
- Gold, M. K.** (2012). *Debates in the Digital Humanities*. University of Minnesota Press.
- Grandjean, M.** (2016). "A social network analysis of Twitter: Mapping the digital humanities community." *Cogent Arts and Humanities*, 3(1): 1171458.
- Hopcroft, J., Khan, O., Kulis, B., Selman, B.,** (2004). "Tracking evolving communities in large linked networks." *Proceedings of the National Academy of Sciences*, 101: 5249–5253.
- McCarty, W.** (2003). "Humanities Computing." In *Encyclopedia of Library and Information Science*. New York: Marcel Dekker.
- Nyhan, J., and Flinn, A.** (2016). *Computation and the Humanities: Towards an Oral History of Digital Humanities*. Cham: Springer International Publishing.
- Nyhan, J., and Duke-Williams, O.** (2014). "Joint and multi-authored publication patterns in the Digital Humanities." *Literary and Linguistic Computing*, 29(3): 387-399.
- Quan-Haase, A., Martin, K., and McCay-Peet, L.** (2015). "Networks of Digital Humanities Scholars: The Informational and Social Uses and Gratifications of Twitter." *Big Data and Society*, 2(1).
- Schreibman, S., Siemens, R. G., and Unsworth, J. (Eds.)**. (2016). *A New Companion to Digital Humanities*. Chichester, West Sussex, UK: John Wiley and Sons Inc.
- Sugimoto, C. R., Thelwall, M., Larivière, V., Ding, Y., and Milojević, S.** (2013). *Mapping Digital Humanities*. Retrieved from <http://did.ils.indiana.edu/dh/>
- Terras, M., Nyhan, J., and Vanhoutte, E. (Eds.)**. (2013). *Defining Digital Humanities: A Reader*. Farnham, Surrey, England : Burlington, VT: Ashgate Publishing Limited ; Ashgate Publishing Company.
- Waltman, L., and van Eck, N. J.** (2013). "A smart local moving algorithm for large-scale modularity-based community detection." *The European Physical Journal B*, 86(11).
- Wang, X., and Inaba, M.** (2009). "Analyzing Structures and Evolution of Digital Humanities Based on Correspondence Analysis and Co-word Analysis." *Art Research*, 9: 123–134.
- Zuccala, A.,** (2006). Modeling the Invisible College. *Journal of the Association for Information Science and Technology*. 57: 152–168.