

- Gill University & Université de Montréal, 119–121. URL: <<https://dh2017.adho.org/abstracts/103/103.pdf>>.
- Fischer, Frank / Trilcke, Peer (2015): "Introducing dlina Corpus 15.07 (Codename: Sydney)". *dlina blog*, 20 June 2015. URL: <<https://dlina.github.io/Introducing-DLINA-Corpus-15-07-Codename-Sydney/>>.
- Kittel, Christopher / Fischer, Frank (2017): "dramavis (v0.4)". *GitHub*, September 2017. URL: <<https://github.com/lehkost/dramavis>>.
- Jannidis, Fotis / Reger, Isabella / Krug, Markus / Weimer, Lukas / Macharowsky, Luisa / Puppe, Frank (2016): "Comparison of Methods for the Identification of Main Characters in German Novels". *Digital Humanities 2016: Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków*, 578–582.
- Moretti, Franco (2011): *Network Theory, Plot Analysis* (= Literary Lab Pamphlet 2), May 2011. URL: <<http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>>.
- Moretti, Franco (2013): "Operationalizing": or, the function of measurement in modern literary theory (= Literary Lab Pamphlet 6), December 2013. URL: <<https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>>.
- Pfister, Manfred (1997): *Das Drama. Theorie und Analyse*. 9th ed. Munich: Fink.
- Trilcke, Peer (2013): "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft". Ajouri, Philip / Mellmann, Katja / Rauen, Christoph (eds.): *Empirie in der Literaturwissenschaft*. Münster: mentis, 201–247.
- Trilcke, Peer / Fischer, Frank / Göbel, Mathias / Kampkaspar, Dario (2016): "Theatre Plays as 'Small Worlds'? Network Data on the History and Typology of German Drama, 1730–1930". *Digital Humanities 2016. Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków*, 385–387.

Visualising The Digital Humanities Community: A Comparison Study Between Citation Network And Social Network

Jin Gao

jin.gao.13@ucl.ac.uk
University College London, UCL Centre for Digital Humanities, Department of Information Studies, United Kingdom

Julianne Nyhan

j.nyhan@ucl.ac.uk
University College London, UCL Centre for Digital Humanities, Department of Information Studies, United Kingdom

Oliver Duke-Williams

o.duke-williams@ucl.ac.uk
University College London, UCL Centre for Digital Humanities, Department of Information Studies, United Kingdom

Simon Mahony

s.mahony@ucl.ac.uk
University College London, UCL Centre for Digital Humanities, Department of Information Studies, United Kingdom

Introduction

An understanding of the intellectual and social structures of Digital Humanities (DH) has been sought by many scholars; some have pointed to the potential usefulness of quantitative methods in such analyses (McCarty, 2003; Terras et al., 2013). A few existing studies have applied quantitative methodologies to analyse publication, conference and social media data (e.g. Nyhan and Duke-Williams, 2014; Weingart, 2016; Grandjean, 2016). This study not only incorporates such approaches but extends them by integrating new analysis and visualisation methods into the wider study of DH's intellectual and social structures.

In this paper, we will introduce research on the citation and social network of DH that is giving rise to new understandings of the field's community structure; scholarly interactions; disciplinary development; and formal/informal communication channels. The citation network of Author Co-Citation Analysis (ACA) comprises 22,321 cited authors across 52,823 cited references from the three core DH journals, while the social network of Twitter Co-Retweet Analysis comprises 3,160 Twitter users and 5,929,609 tweets. To the best of our knowledge, this study is the first to combine bibliometric and social network methods to visualise and compare the DH communities and to uncover their histories. This research contributes to ongoing discussions and debates about the DH knowledge and community structures (Gold and Klein, 2016).

Data Analysis

Citation network

The Author Co-Citation network study was presented at DH2017; this paper extends this earlier analysis. For reasons of clarity, we here give a brief overview of this research. The network has been constructed by collecting the citation data of three core DH journals up to December 2016: *Computers and the Humanities (CHum)*, *Digital Scholarship in the Humanities (DSH)* and *Digital Humanities Quarterly (DHQ)*. 2,582 articles with 52,823 cited references were collected (see Figure 1).

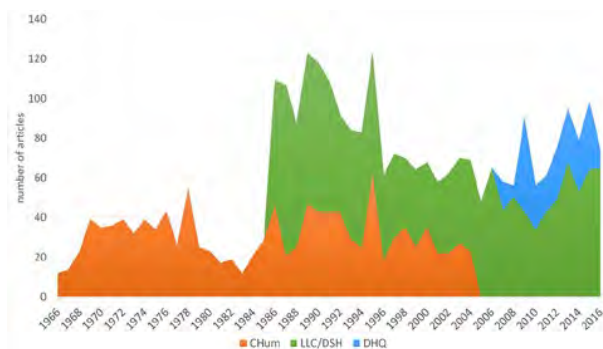


Figure 1. number of articles collected each year (1966-2016)

By using *fractional non-self-citation count* and *exclusive co-citation count* (Zhao and Strotmann, 2008), the weights of nodes and links respectively were calculated for visualisation using the software VOSviewer 1.6.7 (van Eck and Waltman, 2010). An author name disambiguation method (Strotmann et al., 2009) was used, and 22,321 unique cited authors identified. Where possible, other pertinent information was collected (i.e. author full names, gender, country of affiliation, etc.). After counting the occurrences of two authors being cited together, ACA shows the DH intellectual structure and influential topics and scholar groups (Figure 2).



Figure 2. DH Author Co-Citation network

Twitter network

Given DH's early adoption and active use of Twitter (Ross et al., 2011), previous studies have explored the field's scholarly communications and community on Twitter (e.g. Quan-Haase et al., 2015; Grandjean, 2016). This study introduces a new approach (*co-retweet*) to visualise the DH social network.

We have selected all the Twitter users that are followed by the Alliance of Digital Humanities Organisations (ADHO) and its member organisations' (see <http://adho.org/>) Twitter accounts. As dynamic and interdisciplinary as the DH community, it is often difficult and subjective

to select the users by their account descriptions. In contrast, the following relationships by the DH organisations indicate more genuine and representative identification. A total of 3,160 unique users have been collected along with all the 6 million tweets from 21/03/2006 when Twitter was created up to 22:00 (GMT) on 5/11/2017 (see Figure 3).

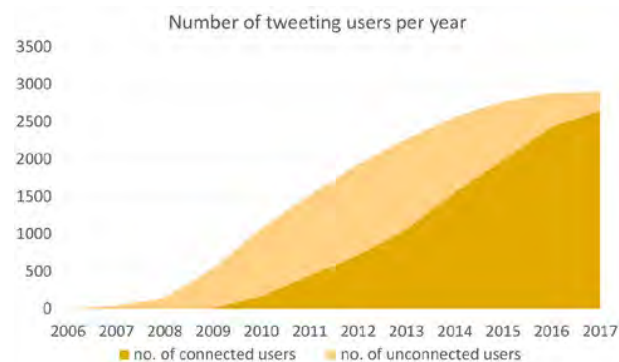


Figure 3. number of tweeting users collected each year

Similar to the citation network method, the Twitter user Co-Retweet network has been constructed by calculating the number of non-self-retweets the user received (*non-self-retweet count*), and the number of same tweets that two users both retweeted (*co-retweet count*) for the weights of the nodes and links respectively. The network resulting visualisations are shown on Figure 4.



Figure 4. DH Twitter Co-Retweet network

Results and comparison

In the citation network, the authors identified distinct topic-based clusters of researchers with backgrounds in information studies and historical literature; in linguistics; in statistical text analysis; in early concordance projects; and biotech influenced text analysis. In contrast, the co-retweet network exhibits grouping based on language and region, with clusters related to scholars in

North America; in Australia; in the UK; and clusters with Francophonic, Germanophonic and Hispanophonic backgrounds.

The Twitter clusters are connected closely whereas clusters on the citation network are more loosely linked. This makes sense, as topics of study are generally more specific and less likely to change, whereas users on social media probably share a wider range of interests. The citation network is based on formal communications and it would take years to get sufficient citations to form links between two scholars. The Twitter network, however, is constructed by more informal interactions between users, and once two users retweeted the same tweet, they immediately build a link on the network.

By visualising both networks during different time periods, this study will also present the topics, disciplines, countries that are involved, and how the networks have been developed and formed over time.

As shown in Figure 5, we divided the 51 years (1966-2016) of historical citation data into five different periods (Hockey, 2004) for visualisation. The citation clusters experienced isolation (1966-1970); connection (1971-1985); consolidation (1986-1990); sub-fields development (1991-2005); and new specialties expansion (2006-2016). Over time, the most cited topics moved from concordance construction, to computational linguistics, then to information and historical literature studies.

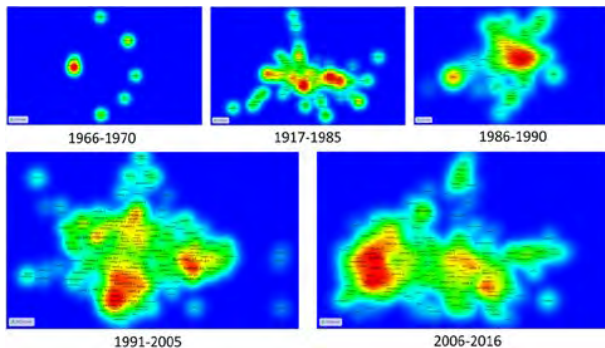


Figure 5. DH Author Co-Citation networks in five periods

As shown in Figure 6, DH Twitter users started to have co-retweet connections in 2009; and then they experienced the beginning of connection (2010); multi-region connection (2010); Anglophonic cluster to centre (2011); Francophonic cluster to develop (2012); North America and UK to separate (2013); Germanophonic to come out (2014); Australian cluster to show (2015); Hispanophonic cluster to emerge (2016); Density continue to move to North America cluster (2017). Over time, the network visualisations show that the density is moving from European clusters towards the North American cluster.

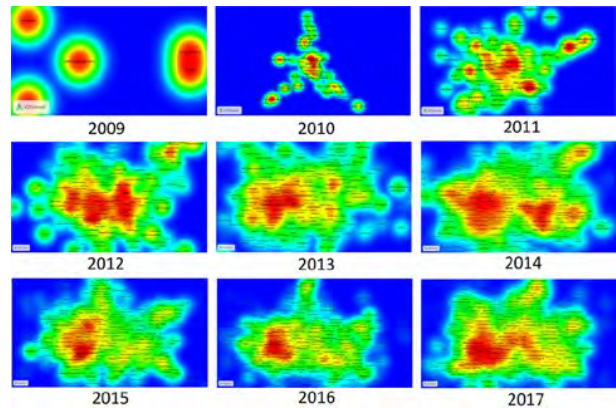


Figure 6. DH Twitter Co-Retweet networks in different years

Discussion and conclusion

This study is not only the first to contribute to the DH history and community studies by visualising and comparing bibliometric and social networks, but also introduces new network approaches (*co-retweet*) to study communications on social media that could support wider social network and data visualisation studies.

As we will discuss, network studies offer powerful but partial ways of studying the aspects of communities that are amenable to quantitative methods. We do not present the visualisations included in this paper as normative representations of the DH "community" or "communities". Nevertheless, when used with caution, network studies can shed new light on important aspects of the historical formation of DH.

There are methodological limitations exist. For example, because the research subjects (cited authors and retweeting users) are not the same group of people (although with much overlap), obvious differences are expected. Besides, the citation lag time has been considered. Other practical methods to identify and study the DH Twitter communities can also be applied.

References

- Davis, L.S., Johns, S.A., Aggarwal, J.K. (1979). Texture Analysis Using Generalized Co-Occurrence Matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1*, 251–259. <https://doi.org/10.1109/TPAMI.1979.4766921>
- Gold, M.K. and Klein, L.F. (2016). *Debates in the digital humanities: 2016*. University of Minnesota Press, Minneapolis London.
- Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*,3:1171458. <https://doi.org/10.1080/23311983.2016.1171458>
- Hockey, S. (2004). The History of Humanities Computing, in: Schreiban, S., Siemens, R., Unsworth, J. (2004).

- A Companion to Digital Humanities*. Blackwell Publishing Ltd, Malden, MA, USA, pp. 1–19.
- McCarty, W. (2003). Humanities Computing, in: *Encyclopedia of Library and Information Science*. Marcel Dekker, New York.
- Nyhan, J. and Duke-Williams, O. (2014). Joint and multi-authored publication patterns in the Digital Humanities. *Literary and Linguistic Computing*,29:387–399. <https://doi.org/10.1093/lilc/fqu018>
- Quan-Haase, A., Martin, K., McCay-Peet, L. (2015). Networks of Digital Humanities Scholars: The Informational and Social Uses and Gratifications of Twitter. *Big Data & Society*,2. <https://doi.org/10.1177/2053951715589417>
- Ross, C., Terras, M., Warwick, C., Welsh, A. (2011). Enabled Backchannel: Conference Twitter Use by Digital Humanists. *Journal of Documentation*,67:214–237. <https://doi.org/10.1108/00220411111109449>
- Strotmann, A., Zhao, D., Bubela, T., (2009). Author name disambiguation for collaboration network analysis and visualization. *Proceedings of the American Society for Information Science and Technology*,46:1–20. <https://doi.org/10.1002/meet.2009.1450460218>
- Terras, M.M., Nyhan, J., Vanhoutte, E. (2013). *Defining Digital Humanities: A Reader*. Ashgate Publishing Company, Farnham, Surrey, England: Burlington, VT.
- van Eck, N.J. and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*,84:523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Weingart, S.B., (2016). dh quantified: A Review of Quantitative Analyses of the Digital Humanities. *the scottbot irregular: data are everywhen*.
- Zhao, D. and Strotmann, A. (2008). All-author vs. first-author co-citation analysis of the Information Science field using Scopus. *Proceedings of the American Society for Information Science and Technology*, 44:1–12. <https://doi.org/10.1002/meet.1450440262>

SciFiQ and “Twinkle, Twinkle”: A Computational Approach to Creating “the Perfect Science Fiction Story”

Adam Hammond

adam.hammond@utoronto.ca
University of Toronto, Canada

Julian Brooke

julian.brooke@gmail.com
Thomson Reuters, Canada

Introduction

In Fall 2015, we were approached by author (and prominent DH-skeptic [2012]) Stephen Marche with a request that we help him use computational textual analysis to allow him to write “the perfect science fiction story.” His

specific request was for a set of “rules” to follow in composing such a story. On consultation with Marche, we devised an approach that would derive “rules” related to theme (using topic modelling) and style (using a variety of techniques, including our own original work on quantifying style) from a hand-selected corpus of Marche’s 50 favourite science fiction stories. In Fall 2016, we sent Marche a list of 14 thematic rules and created a system to provide real-time stylistic feedback on his efforts to meet a set of 24 stylistic targets. In December 2017, the resulting story, “Twinkle Twinkle,” was published in the popular technology magazine *Wired* alongside a set of detailed footnotes explaining and reflecting upon the process of its creation. Whereas Marche and his editors viewed the rule-creating process as “objective” and the publishing professionals interviewed in the piece complained that the resulting story lacked “humanity,” the process in fact blended computational analysis and human interpretation. We regard it as an instance of computer-assisted “creative deformation” rather than “robotic authorship.”

Thematic rules

Below is a selection of the 14 rules we sent to Marche:

1. The story should be set on a planet other than earth.
2. The story should thus NOT be set in space itself.
3. On this planet, there should be an existing, non-human civilization. This civilization should have a hierarchical social structure with a powerful ruler. Inhabitants of this alien civilization should be given clearly non-human names. The protagonists of the story should be humans who are directly observing this civilization from a certain distance and do not consider themselves part of it.
4. The story should be set in a city. The protagonists should be seeing this city for the first time and should be impressed and dazzled by its scale.
5. Part of the action should unfold at night during an intense storm.
6. Include a pivotal scene in which a group of people escape from a building at night at high speed in a high-tech vehicle made of metal and glass that is large enough to live in for an extended period (it should have a bed in it, for instance).
9. Include a scene set on a traditional earth farm, with apple trees and/or corn fields. In this scene, a mother and father are present. (Given the other rules, this is most likely a flashback to a protagonist’s childhood, but I leave the details to you, of course.)
10. Include extended descriptions of intense physical sensations and name the bodily organs that perceive these sensations.
13. DO NOT focus on conventional domestic family life. Marriage should not be a theme. No scenes should depict a conventional bourgeois family (especially a happy bourgeois family) at the dinner table.