

Visualising the intellectual and social structures of Digital Humanities using an invisible college model

Jin Gao

Thesis submitted for the degree of

Doctor of Philosophy

Department of Information Studies

University College London

London, WC1E 6BT

March 2021

Declaration

I, Jin Gao, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Name: Jin Gao

Date: 01 November 2020

Abstract

This thesis explores the intellectual and social structures of an emerging field, Digital Humanities (DH). After around 70 years of development, DH claims to differentiate itself from the traditional Humanities for its inclusiveness, diversity, and collaboration. However, the ‘big tent’ concept not only limits our understandings of its research structure, but also results in a lack of empirical review and sustainable support. Under this umbrella, whether there are merely fragmented topics, or a consolidated knowledge system is still unknown.

This study seeks to answer three research questions:

- a) *Subject*: What research topics is the DH subject composed of?
- b) *Scholar*: Who has contributed to the development of DH?
- c) *Environment*: How diverse are the backgrounds of DH scholars?

The Invisible College research model is refined and applied as the methodological framework that produces four visualised networks. As the results show, DH currently contributes more towards the general historical literacy and information science, while longitudinally, it was heavily involved in computational linguistics. Humanistic topics are more popular and central, while technical topics are relatively peripheral and have stronger connections with non-Anglophone communities. DH social networks are at the early stages of development, and the formation is heavily influenced by non-academic and non-intellectual factors, e.g., language, working country, and informal relationships. Although male scholars have dominated the field, female scholars have encouraged more communication and built more collaborations. Despite the growing appeals for more diversity, the level of international collaboration in DH is more extensive than in many other disciplines.

These findings can help us gain new understandings on the central and critical questions about DH. To the best of the candidate’s knowledge, this study is the first to investigate the formal and informal structures in DH with a well-grounded research model.

Impact Statement

This PhD study has the potential to benefit relevant DH studies as well as science studies of other disciplines for its network visualisations, large-scale dataset, and refined methodology. It could also provide benefits to the development of publishing and social media industry.

Firstly, the main results of citation and Twitter networks provide visual images of the discipline (both as a whole and at individual level) that can be used as teaching tools by scholars and students to gain new knowledge for the ongoing DH debates, and offer a new perspective to revisit its publication and its social communities.

Secondly, it collects bibliometric data from the three most important DH journals that only have limited data indexed in the general citation databases. This study has complemented these databases by adding missing articles and new data (e.g., author gender by name and affiliated country) and building a more comprehensive dataset that will be released open access (potentially with a launch event to promote usage). It saves the duplicated effort of data collection and is stored as a CSV file that can be easily edited and extended.

Thirdly, the refined methodology used in this study can also be replicated by other projects. This study not only adopts the idea of the Invisible College to study the DH community structure, but also extends its model procedure with new and robust approaches. Because the Invisible College model was originally proposed in 2009 when scholarly communications were mainly via traditional ways (e.g., emails and meetings), some of its original methods have now become limited. This study enables this research model to go beyond the traditional disciplinary research system with dynamic types of data, updated methods, and cohesive procedures. In this way, the new Invisible College research model becomes more compatible with the current scholarly communications and more flexible for other disciplines to study their structures and histories.

The output of this study includes an interactive webpage of DH scholar networks that enables the audience to search and filter scholar names within the citation and Twitter networks. This output can serve as a research tool for people who are interested in

the DH community to study individual cases from both bibliometric and social perspectives. The online interactive networks have the potential to be extended to other projects, such as DH genealogy study and DH historical archive study.

Furthermore, this study also provides benefits to improve the indexing standards of academic publishing and to potentially enhance the metadata structure that can serve different purposes. It also offers new insights to the scholarly need for Twitter that could help social media industry to improve the diversity, representation, and inclusiveness of their content.

Table of Contents

Declaration	i
Abstract.....	ii
Impact Statement	iii
Table of Contents	v
List of Tables.....	ix
List of Figures	xi
List of Abbreviations	xvi
1 Introduction	1
1.1 Background context.....	1
1.1.1 Definitional problem	1
1.1.2 Critiques of disciplinary status.....	3
1.1.3 'Digital' and 'Humanities' question.....	4
1.1.4 Lack of DH history	6
1.2 Research questions	8
1.2.1 Subject.....	8
1.2.2 Scholar	11
1.2.3 Environment	14
1.3 Methodological framework.....	16
1.3.1 Invisible college	17
1.3.2 DH invisible college.....	20
1.3.3 Chapter summary.....	24
2 Literature Review	27
2.1 DH subject specialty	28
2.1.1 Journal	28
2.1.2 Conference.....	36
2.1.3 Twitter.....	39
2.1.4 Blog.....	41
2.1.5 Others.....	44

2.2 DH social actors	47
2.2.1 Co-authorship	49
2.2.2 Twitter activity	52
2.3 DH information use environment (IUE)	60
2.3.1 Gender	60
2.3.2 Affiliated country	63
2.4 Discussion.....	68
3 Methodology	70
3.1 Data Sources.....	70
3.1.1 Data from journals	72
3.1.2 Data from Twitter.....	76
3.2 Subject Specialty.....	78
3.2.1 Author co-citation network	78
3.2.2 Twitter hashtag co-occurrence network.....	82
3.3 Social Actors.....	84
3.3.1 Co-authorship network	85
3.3.2 Twitter co-retweet network.....	87
3.4 Information Use Environment (IUE).....	89
3.4.1 Gender	90
3.4.2 Language and affiliated country	93
3.5 Summary.....	96
4 DH Bibliometric Network Analysis	97
4.1 Data collection and cleaning.....	98
4.1.1 Source selection and data extraction	98
4.1.2 Data Cleaning and Formatting	100
4.1.3 Further data collection	105
4.2 ACA network.....	106
4.2.1 Node	106
4.2.2 Edge	109
4.2.3 Network Visualisation	110

4.2.4	Centrality.....	114
4.2.5	Five Periods	116
4.2.6	Discussion and analysis.....	119
4.2.6.1	The ACA network structure.....	120
4.2.6.2	Cluster A – General historical literacy and information science.....	122
4.2.6.3	Cluster B – Computational linguistics.....	128
4.2.6.4	Cluster C – English studies.....	132
4.2.6.5	Cluster D – Early pioneers	137
4.3	Co-authorship network	144
4.3.1	Node	146
4.3.2	Edge	148
4.3.3	Network visualisation	149
4.3.4	Centrality.....	150
4.3.5	Longitudinal analysis	151
4.3.6	Diversity networks	153
4.3.7	Discussion and analysis	154
4.3.7.1	Historical periods	154
4.3.7.2	Gender	163
4.3.7.3	Country.....	169
5	DH Twitter Network Analysis.....	185
5.1	Data Collection and Cleaning.....	186
5.1.1	The dataset	190
5.1.2	Further data collection	194
5.2	Hashtag co-occurrence network	199
5.2.1	Node	199
5.2.2	Edge	202
5.2.3	Network visualisation	203
5.2.4	Centrality.....	204
5.2.5	Longitudinal analysis	204
5.2.6	Discussion and analysis	204
5.2.6.1	Section A – DH and events.....	206
5.2.6.2	Section B – Methods.....	216
5.2.6.3	Section C – Non-Anglophone DH	220
5.2.6.4	Section D – History and other humanities fields.....	225

5.3 Co-retweet network.....	233
5.3.1 Node.....	233
5.3.2 Edge	236
5.3.3 Network visualisation.....	237
5.3.4 Centrality.....	240
5.3.5 Longitudinal and diversity analysis.....	240
5.3.6 Discussion and analysis.....	240
5.3.6.1 Country.....	241
5.3.6.2 Historical periods	254
5.3.6.3 Gender	265
6 Conclusion	272
6.1 Q1 – DH subject.....	273
6.2 Q2 – DH scholar.....	276
6.3 Q3 – DH environment.....	280
7 Reflection and future study	284
7.1.1 Importance.....	284
7.1.2 Limitations	286
7.1.2.1 Dataset.....	286
7.1.2.2 Network visualisation.....	290
7.1.3 Future Study.....	291
8 Appendix.....	297
9 Bibliography	298

List of Tables

Table 4.1: The top 50 cited authors by the fractional nonself citation value, along with total citation and nonself citation counts, data from <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.	108
Table 4.2: Author co-citation matrix of the top 8 cited authors, data from <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017	109
Table 4.3: The top ranked 60 authors whose works published during 1880-1950 were cited the most.	142
Table 4.4: The top 50 authors by the number of publications in journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017	148
Table 4.5: The top 10 authors by the value of Betweenness Centrality on the co-authorship network, data extracted from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.	151
Table 4.6: The relative percentages of the connected authors during different period on the co-authorship networks, data extracted from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.	152
Table 4.7: The top 20 authors ranked by the highest <i>betweenness centrality</i> with gender information (F – female, M – male, U – unknown).	166
Table 4.8: The top 15 countries with the most scholars ranked by the international co-authored rate, data extracted from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.....	175
Table 4.9: The top 20 most frequently co-authored international country pairs, data extracted from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.....	177
Table 4.10: The top 20 domestically co-authored countries, data extracted from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.	177
Table 4.11: The top 15 countries with the most scholars ranked by the betweenness centrality, data extracted from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017.....	180
Table 5.1: The selected organisation accounts and the number of users they followed at the time of data collection.	189
Table 5.2: The top 30 hashtags ranked by its total occurrences.	201
Table 5.3: Hashtag co-occurrence matrix of the top 7 tweeted hashtags with colour scale.	202
Table 5.4: Number of tweets with particular hashtag that included @ symbol and external link in the current dataset.....	213
Table 5.5: Types of the top 100 used Twitter hashtags in the current dataset	227
Table 5.6: The top 20 users who received the most non-self retweets during all years (2009-2017) and in individual years.	236

Table 5.7: The 20 most retweeted tweet by selected users in the current dataset	246
Table 5.8: 10 countries with most users ranked by the value of average <i>betweenness centrality</i> on the co-retweet network. Table also includes average node weight and average number of tweets, data extracted from Twitter, 2006-2017	248
Table 5.9: 10 countries with most users ranked by the value of average <i>betweenness centrality</i> on the co-retweet network. Table also includes average number of unique hashtags used, data extracted from Twitter, 2006-2017	253
Table 5.10: the average value of <i>betweenness centrality</i> , node weight, and tweets posted in gender on the co-retweet network, data extracted from Twitter, 2006-2017	266
Table 5.11: The 30 most used hashtags by DH female and male users in the current dataset, data extracted from Twitter, 2006-2017	268
Table 5.12: Six hashtags as examples used by DH female and male users in the current dataset, data extracted from Twitter, 2006-2017	270

List of Figures

Figure 1.1: The later revision of the knowledge map of Humanities Computing – ‘Mapping the Field to ALLC’ by (McCarty, 2003a). Use in this thesis has been permitted by the authors.....	9
Figure 1.2: A graphic representation of the digital humanities stack (Berry and Fagerjord, 2017, p. 25). Use in this thesis is under the Open Access CC 4.0 licence.	10
Figure 1.3: The original ‘invisible college’ research model (Zuccala, 2006) that demonstrates the organisational structure of the invisible college research model highlighting the interrelationship among its three components.	19
Figure 2.1: The co-occurrence network of high frequency words in 2005 – 2009 by (Wang and Inaba, 2009b, pp. 18–22), and node size represents degree centrality. Use in this thesis has been permitted by the authors.	30
Figure 2.2: Contextual exploration of Digital Humanities (Salah et al., 2015, p. 83).....	32
Figure 2.3: Visualisation of modularity classes in author bibliographic coupling network (Tang et al., 2017, p. 1001), and colours indicate different clusters that match the numbers of the same colour. Use in this thesis has been permitted by the authors.....	34
Figure 2.4: Visualisation of modularity classes in document co-citation network by (Tang et al., 2017, p. 988), and colours indicate different clusters that match the numbers of the same colour. Use in this thesis has been permitted by the authors.	35
Figure 2.5: HASTAC density map of co-occurrence network on humanities-related terms (Puschmann and Bastos, 2015). The image used in this thesis is under Open Access CC 4.0 licence.....	43
Figure 2.6: The screenshot of the topic network visualisation based on data from the Humanist discussion mailing list (1987 - 2014) (McClure, 2014).	46
Figure 2.7: RezoViz view of the scholar co-occurrence network based on the Humanist Discussion Group listserv archives (Rockwell and Sinclair, 2016).....	48
Figure 2.8: Network visualisation of interactions on Twitter during #DH2014 by (Grandjean, 2014; Rochat, 2014). Use in this thesis has been permitted by the author.	55
Figure 2.9: Digital humanities network on Twitter: 2,538 users following each other colour-coded with language. Use in this thesis is under Open Access CC 4.0 licence.	58
Figure 2.10: Physical centres in DH across the globe by (Terras, 2012a).....	64
Figure 2.11: Location of submissions to DH2013, DH2014, and DH2015 (Weingart, 2014c). Use in this thesis has been permitted by the author.	65
Figure 2.12: Co-author network for the journal Digital Humanities Quarterly (2007 – 2014) by (De la Cruz et al., 2015). Use in this thesis is under Open Access CC 4.0 licence	66

Figure 4.1: Area graph – the number of articles collected/published each year in journal <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> (1966-2017)	103
Figure 4.2: Area graph – the number of citations collected/published each year from <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> (1966-2017)	104
Figure 4.3: Area graph – the average number of citations per article collected/published each year for <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i>	105
Figure 4.4: Author co-citation network in DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017	114
Figure 4.5: Author co-citation network of DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-1970	117
Figure 4.6: Author co-citation network of DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1971-1985	117
Figure 4.7: Author co-citation network of DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1986-1990	118
Figure 4.8: Author co-citation network of DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1991-2005	118
Figure 4.9: Author co-citation network of DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 2006-2017	119
Figure 4.10: Density view of author co-citation network in DH, data from journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> , 1966-2017	121
Figure 4.11: The number of cited references that were originally published between 1600 to 1950 in the current dataset	140
Figure 4.12: The number of single-authored and multiple-authored articles collected/published each year in the journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> (1966-2017) ..	145
Figure 4.13: The relative percentage of multiple-authored articles each year in the collected dataset from the journals <i>CHum</i> , <i>LLC/DSH</i> , and <i>DHQ</i> (1966-2017)	146
Figure 4.14: The co-authorship network with all the authors	149
Figure 4.15: DH co-authorship network, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017), graph created by VOSviewer	150
Figure 4.16: DH co-authorship network with average year information, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017), graph created by VOSviewer	153
Figure 4.17: DH co-authorship network of all authors with average year information, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017), graph created by VOSviewer....	161
Figure 4.18: The annual percentages of unique number of authors by gender, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017).....	164

Figure 4.19: The co-authorship network with gender information, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017), graph created by VOSviewer.....	165
Figure 4.20: The number of unique authors in each affiliated country (top 15), data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017).....	170
Figure 4.21: The annual percentages of the international co-authored articles, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017).....	171
Figure 4.22: The annual percentages of the international co-authored articles in DH combined with the percentages of other disciplines 1980 – 2013 in Henriksen’s study (Henriksen, 2016).	172
Figure 4.23: The different author distributions in each affiliated country (top 15), data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017).....	174
Figure 4.24: The co-authorship network with affiliated country information, data from journals <i>CHum</i> , <i>DSH/LLC</i> and <i>DHQ</i> (1966 – 2017), graph created by VOSviewer	179
Figure 5.1: The number of tweeting DH users collected every year on Twitter.	191
Figure 5.2: The number of tweets (including original tweets, retweets, quote retweets, and replies, etc.) from identifiable DH users collected each year.	191
Figure 5.3: The relative percentage of tweets that contained at least one @ sign (or mentioned other users) in the current dataset.	193
Figure 5.4: The relative percentage of tweets that contained URL in the current dataset. .	193
Figure 5.5: Bar chart of the regional distribution of collected tweeting users.....	195
Figure 5.6: Bar chart of the interface language distribution of collected tweeting users	196
Figure 5.7: Top 5,000 words wordcloud based on twitter bios of selected users, constructed using wordle.net.....	197
Figure 5.8: Number of female, male and unknown DH Twitter users each year.....	198
Figure 5.9: The number of unique hashtags used per year.	200
Figure 5.10: Hashtag co-occurrence network in DH, data extracted from Twitter, 2006-2017.	204
Figure 5.11: Hashtag co-occurrence network in DH, data extracted from Twitter, 2006-2017.	205
Figure 5.12: Bottom right part of the hashtag co-occurrence network in DH, data extracted from Twitter.	207
Figure 5.13: The DH hashtag co-occurrence network in 2009, data extracted from Twitter, 2006-2017.....	208
Figure 5.14: The DH hashtag co-occurrence network in 2010, data extracted from Twitter, 2006-2017.....	209

Figure 5.15: The DH hashtag co-occurrence network in 2011, data extracted from Twitter, 2006-2017.....	209
Figure 5.16: The DH hashtag co-occurrence network in 2012, data extracted from Twitter, 2006-2017.....	210
Figure 5.17: The DH hashtag co-occurrence network in 2013, data extracted from Twitter, 2006-2017.....	210
Figure 5.18: The DH hashtag co-occurrence network in 2014, data extracted from Twitter, 2006-2017.....	210
Figure 5.19: The DH hashtag co-occurrence network in 2015, data extracted from Twitter, 2006-2017.....	211
Figure 5.20: The DH hashtag co-occurrence network in 2016, data extracted from Twitter, 2006-2017.....	211
Figure 5.21: The DH hashtag co-occurrence network in 2017, data extracted from Twitter, 2006-2017.....	212
Figure 5.22: Bottom left part of the hashtag co-occurrence network in DH, data extracted from Twitter.....	216
Figure 5.23: Left part of the hashtag co-occurrence network in DH, data extracted from Twitter.....	221
Figure 5.24: Top right part of the hashtag co-occurrence network in DH, data extracted from Twitter.....	226
Figure 5.25: Line graph of the relative percentages of history-related hashtags' annual occurrences in the current dataset (2009-2017).....	230
Figure 5.26: Line graph of the relative percentages of education-related hashtags' annual occurrences in the current dataset (2009-2017).....	231
Figure 5.27: Line graph of the relative percentages of politics-related hashtags' annual occurrences in the current dataset (2009-2017).....	231
Figure 5.28: Line graph of number of users against number of non-self retweets received	234
Figure 5.29: Bar chart of number of users against number range of non-self retweets received	235
Figure 5.30: User co-retweet network in DH, data extracted from Twitter.	238
Figure 5.31: User co-retweet network in DH colour-coded with different languages, data extracted from Twitter.	242
Figure 5.32: User co-retweet network in DH colour-coded with different affiliated countries, data extracted from Twitter.	243
Figure 5.33: User co-retweet network in DH colour-coded with different time zones in the USA and Canada, data extracted from Twitter.	244

Figure 5.34: User co-retweet network in DH 2009 colour-coded with different affiliated countries, data extracted from Twitter.....	255
Figure 5.35: User co-retweet network in DH 2010 colour-coded with different affiliated countries, data extracted from Twitter.....	256
Figure 5.36: User co-retweet network in DH 2011 colour-coded with different affiliated countries, data extracted from Twitter.....	257
Figure 5.37: User co-retweet network in DH 2012 colour-coded with different affiliated countries, data extracted from Twitter.....	258
Figure 5.38: User co-retweet network in DH 2013 colour-coded with different affiliated countries, data extracted from Twitter.....	259
Figure 5.39: User co-retweet network in DH 2014 colour-coded with different affiliated countries, data extracted from Twitter.....	260
Figure 5.40: User co-retweet network in DH 2015 colour-coded with different affiliated countries, data extracted from Twitter.....	261
Figure 5.41: User co-retweet network in DH 2016 colour-coded with different affiliated countries, data extracted from Twitter.....	262
Figure 5.42: User co-retweet network in DH 2017 colour-coded with different affiliated countries, data extracted from Twitter.....	263
Figure 5.43: DH Twitter co-retweet networks in different years	265
Figure 5.44: The gender distribution in the top 10 countries with the most users in the current dataset, data extracted from Twitter, 2006-2017	267

List of Abbreviations

A&HCI - The Arts & Humanities Citation Index

AAAG - the Annals of the Association of American Geographers

ADHO - The Alliance of Digital Humanities Organizations

ALH - American Literary History

API - Application Programming Interface

CHum - Computers and the Humanities

DADH - The International Conference of Digital Archives and Digital Humanities

DH - Digital humanities

DHd - The German Digital Humanities Conference

DHQ - Digital Humanities Quarterly

DSH - Digital Scholarship in the Humanities

GLAM - Galleries, Libraries, Archives, and Museums

HASTAC - The Humanities, Arts, Science, and Technology Alliance and Collaboratory

IUE - Information Use Environment

JCR - Journal Citations Report

LLC - Literary and Linguistic Computing

MLA - The Modern Language Association

THATCamp - The Humanities and Technology Camp

1 Introduction

1.1 Background context

What is Digital Humanities (DH)?¹ Depending on who you ask, the answers may vary significantly, because different people recognise DH differently. Starting from its origins in concordances and tool-building projects (Nyhan and Passarotti, 2019), the current rubric of DH includes a variety of topics, such as linguistics, lexicography, literary studies, history, art history, classics, archaeology, music, performing arts, philosophy, religion, videogame, image processing, and many more (Schreibman et al., 2004a, 2016).

The term 'big tent' has been used to describe this diverse and inclusive scope of DH (Svensson, 2012), but such a term results in challenges to clearly examine its boundaries and structure. Still, for both practical and pragmatic purposes, answering the question of 'what is DH' is needed. It not only helps us gain a better understanding of 'who we are' but also assists in maintaining a healthier community and a more sustainable development.

This section (1.1 Background context) briefly reviews previous efforts that tried to answer such questions from four perspectives, i.e., DH definition, disciplinary status, the relationship between its 'digital' and 'humanities' components, and its history study. By summarising fundamental issues and the state-of-art development in DH, this section provides a research context. Further questions are raised at the end of each discussion in order to introduce the problems that have not yet been resolved, and thus, to form the research questions of the current study in section 1.2 that follows.

1.1.1 Definitional problem

The field that we call 'Digital Humanities' today is often traced back to 1949 and the work of Roberto Busa (Nyhan and Passarotti, 2019). In the following years, the field went under various names, for example, 'Humanities Computing', 'eHumanities', 'Digital Resources in the Humanities', 'Literary and Linguistic Computing', and

¹ This study has followed the previous convention that describes Digital Humanities as a singular collective noun, e.g., (Liu, 2016, p. 1546; Berry and Fagerjord, 2017, p. 10).

‘Humanist Informatics’ (Nyhan et al., 2013, p. 2). Distinctions exist among these different terms, particularly between ‘Humanities Computing’ and ‘Digital Humanities’ (Schreibman et al., 2004b; Rockwell et al., 2011, p. 207). Some believed that the former was used more often before 2005 until some ‘specific circumstances’ in 2005 and 2006 that witnessed the rise of the use of ‘Digital Humanities’ (Kirschenbaum, 2010, p. 2). The change from ‘Humanities Computing’ to ‘Digital Humanities’ is sometimes used to evoke the sense of transitioning from ‘support services’ to progressive intellectual efforts with its own professional methods and practices (Hayles, 2012; Vanhoutte, 2013).

Nevertheless, even after the extensive use of ‘Digital Humanities’, new names are still being suggested. Terms such as ‘Digital Liberal Arts’ (Pannapacker, 2013), ‘Digital Studies’ (Stiegler, 2012), and ‘Computational Criticism’ have been proposed to widen the range of its research focus, while ‘Digital Humanities’ was being criticised as ‘means nothing’ (Dinsman, 2016).

Regardless, the term ‘Digital Humanities’ is now very much accepted and used by most practitioners, institutions and organisations. The way it has contributed to the identification of the field and the community formation is significant (Berry, 2011). In this research, to follow the previous convention (Nyhan and Flinn, 2016, p. 2), the term ‘Digital Humanities’, as it is currently known, is used to refer to this discipline (c.1949 - now) throughout this thesis for convenience purposes. Yet, other terms are indicated in some chapters to assist the discussion of historical context and clarity.

However, the settlement of the field’s name does not mean that we have the answer to ‘what is DH’. In order to unpack this question, defining the landscape of DH has become a central concern during the last two decades. Many critical efforts have been made to define DH (e.g., Geoffrey Rockwell et al., 2012b; Terras et al., 2013; Gibbs, 2013). According to different groups of people, DH was ‘a critical investigation and practice of the methods of humanities research in the digital medium’ (Flanders et al., 2007), a ‘term of tactical convenience’ (Taporwiki, 2011), a ‘social category, not an ontological one’ (Alvarado, 2012, p. 50), or a ‘term can mean anything’ (Ramsay, 2013a). Some agreed with McCarty that it is a question not to be answered but continually to be explored and refined (McCarty, 2003a; Terras et al., 2013). Therefore,

while the focus of defining DH seems to move away from the central debates in the past years, it is still a valuable and ongoing work to be done continuously.

To fulfil this need, one should step away from just focusing on nomenclature and definition, and, alternatively, look for open-ended and flexible methods that can reflect new changes and dynamic disciplinary developments. As Siemens noted, defining ‘who we are’ is no longer only about finding ‘what we do’ but also ‘where, how, and with whom it is we do what we do’ (Siemens, 2016, p. xxiv). With this idea in mind, new questions emerged. How can we positively and critically present a DH outlook that is beyond the current context? How can we improve the previous definitional conversations and build upon a new study that witnesses the history of DH from the early stage of a small group of scholars to a large network of practitioners?

1.1.2 Critiques of disciplinary status

While DH continues to be vigorous and rapidly growing, it remains controversial if it should be referred to as a discipline at all (Thaller, 2012, p. 12; Schreibman et al., 2016, p. 479).

Although when referring to an academic territory, studies often use ‘discipline’ and ‘field’ interchangeably (e.g., Favero, 2010, paras. 1–2), there are certain distinctions that exist (Becher and Trowler, 2001a, p. 41). Despite the ongoing challenges and discussions about its own definition (e.g., Gascoigne et al., 2010), discipline, as this candidate understands, refers to a branch of knowledge studied at university level. As Terras mentioned, ‘institutionalising the subject would seem to give gravitas: if you can point at an academic department, the discipline exists’ (Terras, 2010, p. 175). Field, on the other hand, is often specified as a sub-discipline or sub-branch of knowledge, and some of the fields are yet to be recognised by other disciplines (Krishnan, 2009, pp. 4–7).

The question of whether DH is a discipline and what problems this disciplinary status may cause to other fields was asked as early as 1999 (Nyhan and Flinn, 2016, p. 6). From the educational and pragmatic point of view, many believe that DH already had all the qualities of a discipline and its establishment, and therefore, should be treated as a discipline (Svensson, 2009, para. 16). These qualities include a well-established and expanding community, regular journals and conferences, different levels of

teaching programmes, and centres in major universities around the world (Gold, 2012, p. ix; Nyhan and Flinn, 2016, p. 7).

Although these qualities have shaped an academic ‘ecosystem’ with certain maturity to demonstrate that DH is a legitimate discipline (Siemens, 2016, p. 1; Berry and Fagerjord, 2017, p. 13), many critiques have challenged this disciplinary development by raising questions about a variety of ‘ills’ including the lack of diverse representation and political commitment, unbalance between research and teaching, less open publication copyright, and preference for funding-driven projects (Gold, 2012, p. xii). Even the ‘ecosystem’ that may help DH establish its disciplinary status is criticised as ‘ambiguous’ and ‘marginal’ (Poole, 2017, p. 95).

Sustainability of DH projects is also described as the ‘elephant in the room’ that is difficult to maintain in the disciplinary development (McGann, 2010). Ideally, DH centres and projects need long-term maintenance and ongoing efforts, but in fact most of them rely on short-term and one-off funding that makes them unsustainable and ‘uncertain’ (Rockenbach, 2013, p. 6), and according to a survey by Zorich, 78% of the centres suffered at least one unsuccessful partnership because of sustainability issues (Zorich, 2008, p. 34).

Despite the evidence to demonstrate DH as a discipline or to question its disciplinary development, many related questions also require attention. For example, how does one improve the ‘ambiguous’ position of the DH community in academia? Or maybe one should follow McCarty to reject the idea of discipline and propose a concept of interdiscipline which sees DH from a very different perspective (McCarty, 1999)? How to answer the doubts and debates that exist in the feasibilities of the key skills that the ‘discipline’ ‘must have’ (e.g., the ability to code) (Ramsay, 2013b)?

1.1.3 ‘Digital’ and ‘Humanities’ question

DH has emerged at the intersection of the ‘digital’ and the ‘humanities’ and has advantages to gather approaches and knowledge from both sides. However, what is (or should be) the relationship between ‘digital’ and ‘humanities’ is still not clear.

In general, this is a ‘meaning problem’, as Liu discussed, which is related to the balance between numbers and meaning (Liu, 2013, p. 411), between building and

thinking (Sample, 2011), between ‘hack’ and ‘yack’ (Nowviskie, 2016a), making and theorising, and between doing and saying (Cecire, 2011). This problem exists in ‘many parts of the field’ when researchers try to get from quantitative numbers and evidence to meaningful, interpretable and qualitative insights (Liu, 2013, p. 411).

On one hand, some scholars emphasise the merits of ‘making’ that help non-makers understand complex resources, and such a ‘digital’ movement of practicality is not only found in DH but also in a much broader scope that encourages everyone to code, e.g., Code Academy and primary education (Curtis, 2013). Consequently, especially for those new to DH, only the ‘digital’ is seen as methods that are dynamically changing, not the ‘humanities’. However, Siemens argued that the humanistic scholarship has been changing and evolving since its emergence, and it is the ‘humanities’ that is the key element in DH, not the ‘digital’ (Siemens, 2016, p. xxi).

Still, distinctions exist between DH and the humanities, such as different institutional environments (mostly, DH as interdisciplinary centres while the humanities has its own independent infrastructure), working pattern (DH claims to have more collaboration), expenses (DH has more expensive equipment and resources), engagement with people outside of academia (more ‘non-specialists’ seem to be involved in DH works, e.g., crowd-sourcing, although there are ethical discussions ongoing) (Williamson, 2016). Additionally, DH has extended the humanities publishing landscape even further, from journals and books to coding repositories (e.g., GitHub or Bitbucket), interactive visualisations and websites (Gold, 2012, p. xi).

On the other hand, however, ‘thinking’ from the ‘humanities’ side is often assumed to be more clever and is rewarded in academia, while ‘making’ and ‘building’ are not, especially when it comes to evaluation and examination (Ramsay and Rockwell, 2012). Especially in the early years, digital methods were sometimes identified as something done to the humanities. The funding, publishing, and marketing influences reinforced this impression that made the humanities scholars seem ‘under attack’ and start to doubt their key focus of research in this digital age (G Rockwell et al., 2012, para. 29).

Even though efforts were made by DH practitioners to ease the negative impressions, DH was treated as rather a ‘Trojan horse’ to the prospect of the institutional future by the humanities. DH scholars, especially at early stage, were seen as ‘unwelcome

messengers' of the approaching transformation to humanistic research and thinking (Geoffrey Rockwell et al., 2012a, para. 29).

Unlike the 'humanities', the impression of DH in the eyes of the 'digital' seems more amiable. Some computer scientists have paid attention to the applications and cultural changes in the humanities. In addition to the increasing collaborations across different departments, many institutions of computing specialty have hired DH scholars to assist their work and research on digital applications in the humanities, for example the Google Cultural Institute (Google, 2018) and the Alan Turing Institute (The Alan Turing Institute, 2018). Also, many DH research topics, tracks and panels are seen at prestigious computer science conferences, for instance, the WWW (World Wide Web) 2018 conference (WWW2018, 2018) and the International Conference on Web and Social Media (ICWSM19, 2018).

Nevertheless, scholars who are involved in the 'digital' or 'humanities' debates mostly rejected this binary opposition. As Warwick stated, they 'stress commonality rather than widen differences' and 'slide away from the initial provocation and seek a middle ground of agreement' (Warwick, 2015, pp. 540–541; and see, e.g., Cecire, 2012; Grusin, 2013; Ramsay, 2013a; Bond et al., 2017). However, she also pointed out that such an alleged binary debate was an 'inevitable pain of developing into a mature discipline' (Warwick, 2015, p. 538).

Yet, even if binary debates are inevitable, there remains the problem of what proportions do 'digital' and 'humanities' account for respectively? Where are their territories in the landscape of DH?

1.1.4 Lack of DH history

The history of the field that nowadays we call Digital Humanities can be traced back to the year of 1949 when Father Roberto Busa began the pioneering concordance project of the works of St Thomas Aquinas and related authors (Busa, 1950). Although some argue that DH as a field only emerged more recently (Berry and Fagerjord, 2017, p. 10), others claim that it started in the nineteenth century (Brandeis Library, 2013; Hayes, 2017). Busa is mostly accepted as the starting point, and thus the research practice of DH has been around 70 years. The history of DH was almost ignored and unstudied for a long time, and the urgent need for and emphasis on its history only

gained increasing attention in recent years. In 2016, Nyhan and Flinn pointed out that the study of DH history had started to emerge and was absolutely necessary (Nyhan and Flinn, 2016, p. 14). Many believe that the lack of the history of DH has prevented its disciplinary development and possible future improvement (McCarty, 2011, pp. 4–6; Nyhan and Flinn, 2016, p. 15), and McCarty even described the ‘crying need for history’ (McCarty, 2011, p. 6).

Why was it not studied? Various complicating factors may be enumerated. For example, the longer history of the wider humanities also lacks study, and this provides no useful tradition or framework for the study of the history of DH to follow (Nyhan and Flinn, 2016, p. 14). There are obstacles in the process of such historical studies, too. To study DH history, scholars need historical evidence, e.g., archival materials. Although more and more archives are being digitised and shared online, the archives related to DH are very limited, and issues of copyright make the problem worse. Nyhan has discussed the difficulties to detect and collect these materials (Nyhan and Flinn, 2016, p. 11).

Nevertheless, the history of digital humanities has attracted increasing attention in recent years. In addition to archival research, other methods are also being used, e.g., the narratives from the ‘contemporary elders’ (Cronin and Sugimoto, 2014, pp. 372–376), general literature reviews from comprehensive bibliographies (McCarty, 2003b, p. 1224), interviews and oral histories (Nyhan and Flinn, 2016), statistical results (Terras, 2012a), bibliometric analysis (Wang and Inaba, 2009a), and social network analysis (Grandjean, 2016). Even with these various methods, however, few efforts have been made throughout the disciplinary development to discuss what happened in the past (e.g., Burton, 1982; Raben, 1991; McCarty, 2003b; Hockey, 2004; Nyhan and Flinn, 2016), and the number of studies is far from enough.

How could we go beyond these efforts to solve this urgent need for history? Is there an inclusive method to study as many historical figures as possible from a quantitative perspective? Who were the adopters in different periods of DH development? Were they well-known or less-known to the current DH community? What topics was this field involved in during the past and how did they develop over time? How did the DH intellectual map shape over time?

1.2 Research questions

Considering the scale of this PhD study and the capacity of researching and visualising the DH intellectual and social structures, this study begins to organise the questions raised in the above section 1.1 (Background context) into three main aspects: *subject*, *scholar*, and *environment*. The three aspects are intended to cover the questions asked above, and each has one primary question followed by a series of sub-questions. Below is a summary of the research questions.

- a) *Subject*: What research topics is the DH subject composed of?
- b) *Scholar*: Who has contributed to the development of DH?
- c) *Environment*: How diverse are the backgrounds of DH scholars?

1.2.1 Subject

What research topics is the DH subject composed of? Addressing ‘what gets counted as DH’ is not only useful in founding new research and education, but also crucial to our understanding towards the ‘digital’ and ‘humanities’ question, its disciplinary status and problems of definition (Barnett, 2014, pp. 68–69).

As discussed above, different people recognise DH differently, and the ‘recognition or misrecognition’ from different people might be one of the factors that stimulate the debates about what constitutes DH (Barnett, 2014, p. 64). For example, Raben might be one of the first scholars to describe the scope of the field, but during that time in 1966, he did not actually separate the field from the traditional humanities (Raben, 1966, p. 1). Similarly, in searching for the meaning of DH, Liu, a professor in English Studies, saw DH serve as ‘a shadow play for a future form of the humanities’ and ‘much of which affiliates with older humanities disciplines’ (Liu, 2013, p. 409), and he also pointed out the problem of lack of theory (Liu, 2012a).

Some see DH as ‘becoming interdisciplinary’ (McCarty, 2015, p. 79). Together with Harold Short, McCarty presented a graphical representation of the field’s intellectual map (McCarty, 2003b, p. 1225), which he later revised (McCarty, 2003a, p. 119). As shown in Figure 1.1, he illustrated the methodological commons surrounded by various subjects (e.g. text, image, 3D visualisation) that link to many disciplines at the

top and ‘clouds of knowing’ at the bottom. Where all these branches meet is the central area of field with formal methods.

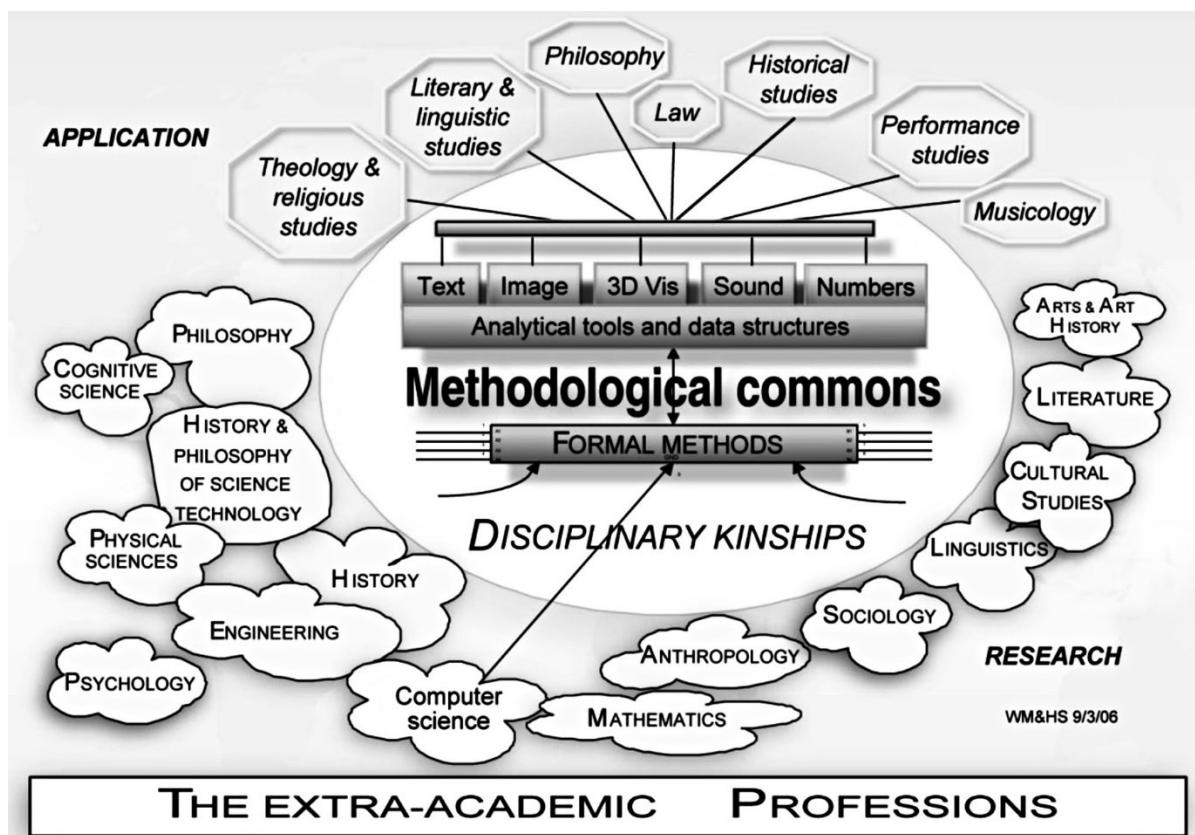


Figure 1.1: The later revision of the knowledge map of Humanities Computing – ‘Mapping the Field to ALLC’ by (McCarty, 2003a). Use in this thesis has been permitted by the authors.

Later McCarty added that ‘[...] digital humanities has a centre all over the disciplinary map and a circumference that is at best uncertain’ (McCarty, 2015, p. 79). Poole seconded him as ‘both the center and the boundaries of digital humanities remain amorphous’ (Poole, 2017, p. 94).

Aiming to discuss the field’s history, Hockey reviewed the scope of the field by highlighting the important events and ‘typical research outputs’ chronologically. From her view, the focus of the field was around ‘computer-assisted lexicographical studies, authorship, stylistic studies, and the limitations of the technology’ (Hockey, 2004, p. 5). We might argue that such a list of topics is subject to researcher’s time, experience, and knowledge.

Svensson examined the landscape of DH by investigating its engagement with information technology. Coming from a linguistic and technology background, he categorised five major modes of engagement between the humanities and information technology in DH, i.e., ‘information technology as a tool, as a study object, as an expressive medium, as an experimental laboratory and as an activist venue’ (Svensson, 2010, para. 101).

Similar to Svensson, by expressing interests in DH particularly in algorithms, software and code, Berry and Fagerjord summarised a broad outlook of DH as focusing on ‘tools’ and computational methods (Berry and Fagerjord, 2017, p. 16). They argued that DH research focus was heavily concentrated on technology and was continually moving away from critical engagement and theoretical research (Berry, 2011). In addition, by borrowing the idea of a ‘stack’ from computer sciences, they illustrated DH as a graphical map – ‘DH stack’ with different layers of practices (Figure 1.2).

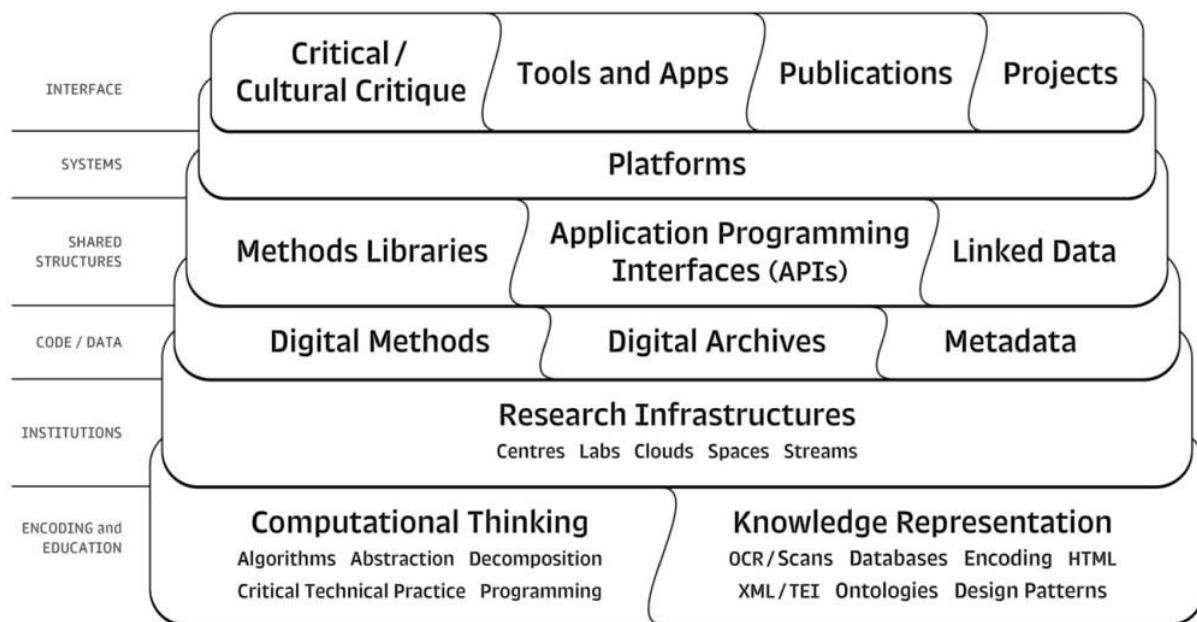


Figure 1.2: A graphic representation of the digital humanities stack (Berry and Fagerjord, 2017, p. 25). Use in this thesis is under the Open Access CC 4.0 licence.

In general, narratives can be complex and contradictory (Nyhan et al., 2015). Studies reviewed above held different conclusions about the DH intellectual structure. Yet, most remained either at the level of general narratives or specific cases, and their different focuses, thus, resulted in different perception and representation of the DH subject.

As different people see DH differently, is there an approach that could assist in finding what constitutes the field of DH from a broader view that can overcome the different recognition or misrecognition from different people? It is necessary to revisit this question that still sits in the centre of DH debates, and applying a new and practical method could help us exceed the existing research framework and deal with this question from a new perspective.

While the methodological framework will be discussed in the next section (1.3 Methodological framework), here we start shaping the first research question by emphasising a need for understanding 'what is DH' from a different perspective that returns to the very topic that is related to the field:

What research topics is the DH subject composed of?

What are the main topics (or subject specialties) under its so-called 'big tent'? How do these topics relate to each other? Have they consolidated as a cohesive disciplinary system or still remained fragmented? Is there an intellectual structure of DH that can be identified? How do they develop over time? What disciplines are involved in the development of DH? Which discipline(s) contribute the most? and during what periods? Will the DH development eventually result in a radical reconfiguration and being absorbed into each of the humanities disciplines (Berry and Fagerjord, 2017, p. 12)?

1.2.2 Scholar

Who is a DH scholar? Who gets counted and who gets excluded? This question has been asked, debated, and criticised extensively and was originated from the 'History and Future of Digital Humanities' panel at the MLA 2011 (The Modern Language Association 2011) (Ramsay, 2013b).

Led by it, a series of questions were asked both online and offline. For example, do we get included if we have been to a DH event or conference (e.g., THATCamp – The Humanities and Technology Camp)? Must we have a membership in a DH organisation (e.g., ADHO – The Alliance of Digital Humanities Organizations) in order to be a DH scholar? Do we have to know how to code to be a DH scholar? Does it have to be about works situated at a DH centre? Do we have to publish in a DH journal or book? Do we need to be on Twitter or use a blog?

These questions frame the values of the field. DH claims to value openness, collaboration, diversity and inclusiveness, and these values help to unite the community to confront challenges (Spiro, 2012). However, this does not mean that its disciplinary boundaries can be unlimited and unfettered. The notion of the ‘big tent’ seemingly helped to solve the debates about ‘who’s in and who’s out’ by encouraging the building of an inclusive community and to collaborate and cooperate. Yet, continuous widening and expanding the territory of DH became problematic. This problem was not only about making the tent too big, but also setting barriers for understanding how the field was ‘epistemologically textured’ (Svensson, 2012). Thus, instead of ‘big tent’, concepts such as ‘no tent’, ‘trading zone’ or ‘meeting place’ were suggested to be more beneficiary to the development of DH (Svensson, 2016, p. 82).

Hence, returning to the question of ‘who’s in and who’s out’, in order to identify and study the DH community, this thesis begins to adjust the question more practically as ‘who has contributed to the development of DH?’². Although these two questions appear very much alike, they have different emphases on the academic impact. The latter pays more attention to the output and influence that a scholar makes towards DH and its community.

However, how do we define ‘contribute’ and what should be counted as ‘the DH development’? Is it just the publications that are normally considered as academic production? Or maybe we should consider an altmetric-like approach that also includes social media contributions? Should conference attendance or organisation membership be counted as academic influence, too? As we can see, continuing to extend this question involves more methodological issues that require pragmatic process and rational design, and it will be further discussed in section 1.3 Methodological framework and chapter 3 Methodology. Nevertheless, from a new aspect, shaping this second research question helps to see the community with more tangible aims and measurable approaches.

Digital technology changes the way scholars communicate and collaborate (Borgman, 2009, p. 9), and DH is no exception (Poole, 2017, pp. 97–98). Firstly, DH was one of

² This study defines the time range of ‘the development of DH’ as its mostly agreed historical period, i.e., c.1949 – now.

the first academic communities that embraced alternative modes of scholarly communication that encouraged participation, sharing, collaboration, openness and community formation (e.g., the Humanist listserv, Blog, Twitter, and other social media) (McCarty, 1992; Ross et al., 2011; Puschmann and Bastos, 2015). On the other hand, these social media channels are informal interactions that not only introduce ‘cocktail-party-like’ casual conversations (Parry, 2011, para. 6) but also raise ‘heated, public, personal and unpleasant’ debates (Warwick, 2015, p. 544). Few studies have been conducted to investigate important, urgent and ‘troubling’ concerns about DH using social media (Poole, 2017, pp. 98–99). For example, does social media strengthen existing hierarchies? Is using social media a sustainable way to conduct research and scholarship? How to ease the ‘chasm yawns between scholars’ awareness of social media and their use of it’ (Ross, 2012, p. 26)?

Secondly, collaboration is valued as the ‘fulcrum’ in DH (Rockenbach, 2013), and its collaborative stereotype is seemingly well-known (Fitzpatrick, 2011; Koh, 2012), but some think it is ‘problematic’ and find that the collaborative character is not unique to DH (Deegan and McCarty, 2012, p. 2; Nyhan and Duke-Williams, 2014a). Collaboration in DH is difficult to build and maintain, because it not only requires a series of commitments from each team member that might come from different specialties (e.g., trust, understanding, consensus, compromise, balance, responsibility, management) (Siemens, 2009), but also faces high probability of ‘running aground’ (Poole, 2017, pp. 104–105). The need for researching on DH collaborative behaviours and practices ‘is as urgent as in the sciences’ but remains far from well-studied (Poole and Garwood, 2018, p. 184).

Investigating the patterns of communication and collaboration in DH not only fulfils the urgent needs mentioned above, but also helps to improve future scholarly interactions among interdisciplinary scholars. Therefore, the second research question is formed as:

Who has contributed to the development of DH?

Who is known or less-known to the DH community? How do they communicate and collaborate? What publishing pattern and social structure can be identified among these scholars? What might be the determining factors of such structures?

1.2.3 Environment

How diverse are the backgrounds of DH scholars? To answer this question, one needs to understand what is meant by 'diversity'.

In sociology, diversity refers to 'the position of a population along a continuum ranging from homogeneity to heterogeneity with respect to one or more qualitative variables' (Lieberson, 1969, p. 851). It is specified to describe the degree of differences in characteristics among different population groups, such as in race, ethnicity, age, gender, religion, physical abilities, sexual orientation. When measuring demographic diversity, an index is often used by calculating the percentage based on any sample to illustrate the probability that two subjects belong to different diversity groups. For example, if all subjects are from the same group, the index is 0%, and if half are from one group and half from another, it is 50%.

In DH, 'core' and 'peripheral' groups both contributed to the development of knowledge (Huggett, 2012). For example, studies have uncovered hidden contributions of women to the field (Nyhan and Terras, 2017). People of different colour, race, country, using different languages also formed dynamic networks of communication and collaboration in DH (Risam, 2015a; Gallon, 2016; Mahony, 2018; Earhart, 2018).

Recent years have witnessed increasing attention to themes, such as diversity, decolonisation, and scholarly environment and identity in DH³. Questions were asked and widely debated by many scholars, such as 'why are the digital humanities so white?' (McPherson, 2012), 'where is cultural criticism in the digital humanities?' (Liu, 2012b), 'can information be unfettered?' (Earhart, 2012), 'can we describe digital archives as feminist?' (Wernimont, 2013), 'can digital humanities mean transformative

³ There were growing numbers of panels at major DH events organised by, e.g., MLA (Modern Language Association), ADHO (Alliance of Digital Humanities Organizations), and HASTAC (Humanities, Arts, Science, and Technology Alliance and Collaboratory), discussing the diverse scholarly backgrounds in DH. A new category of 'Digital Humanities – Diversity' was added to keywords for ADHO conference in 2016 (Weingart, 2015a). Related journal issues and books were also published, such as the *Digital Diversity: Cultures, Languages and Methods* special issue on *LLC (Literary and Linguistic Computing)* (Spence et al., 2013), the *Feminisms and DH* special issue on *DHQ (Digital Humanities Quarterly)* (Wernimont, 2015), edited books, e.g., (Schreibman et al., 2016; Gold and Klein, 2016; Berry and Fagerjord, 2017; Losh and Wernimont, 2019).

critique?' (Lothian and Phillips, 2013), 'what is digital humanities, and why are they saying such terrible things about it?' (Kirschenbaum, 2014). In general, these events, publications and efforts have boosted the conversations about the significant role of diversity and complex environmental background in DH.

Regardless of such growing attentions, it is still astonishing to see conflicts between academic practices and 'structural misogyny and racism' (Losh and Wernimont, 2019, p. ix), and 'the state of gender in digital spaces around the world has only grown more dismal' (Risam, 2015a, para. 1). People who support DH feminism received 'harassment' online and were alleged to be 'toxic' (Risam, 2015b; Massanari, 2017). Such conflicts not only establish barriers, but also show that diversity conversations are not as important as other topics in DH. This demonstrates that diversity studies in DH is necessary.

Along with the global expansion of DH, a series of issues that were previously neglected began to emerge, for instance, a homogeneous group made up of white scholars at the hierarchical centre without understanding across differences (Bailey, 2011), large blank spaces on Terras's *Quantifying the Digital Humanities* map (Terras, 2012a; Fiornonte, 2012; Mahony, 2018), a distinct linguistic divide 'between UK/USA and the rest of the world' and the domination of English as the language of communication (Clavert, 2012) and publication (Fiornonte, 2014) with a distribution heavily focused on countries with high-income economies (O'Donnell et al., 2015). People started to be aware that DH was not 'as open and universal as it had initially perceived itself', and problems related to gender, race, language, and class were discussed more frequently on mainstream DH channels (Galina, 2014). Invested efforts were not recognised by larger communities, practices were blind to certain privilege and exclusion, and many were concerned about the development of the field (Bailey et al., 2016).

These issues mentioned above indicate a still urgent need for studies of diversity and the scholarly environment in DH. Investigating such urgent need provides us with a broad insight into the understanding of the purpose of DH and its community, and the interrelationship and mutual influences between diversity and disciplinary development. By doing so, this study helps to transform the field into a more

sustainable and representative place, as Barnett stated in the ‘Brave Side of Digital Humanities’:

What happens when we shift difference away from a deficit that must be managed and amended (with nods in the direction of diversity) and toward understanding difference as our operating system, our thesis, our inspiration, our goal? From this perspective, highlighting the brave side of digital humanities isn’t an act of transformative resolution, but is about reframing and recognizing which links were already there and which links are yet to be made. (Barnett, 2014)

Therefore, this study explores the relationship between gender, affiliated countries, language and DH, and sheds light on the demography of the DH community and their influences on DH communication and collaboration. The third research question of this study, accordingly, is shaped as:

How diverse are the backgrounds of DH scholars?

What is the gender distribution in certain DH communities? How many affiliated countries are involved? How do we go about finding DH scholars and communities that are not yet connected? To what degree do these diversity factors influence the DH intellectual and social structures?

1.3 Methodological framework

The research questions set out above are formed in three aspects, i.e., *subject*, *scholar*, *environment*. Without a systematic methodological framework and a robust research design to unite them, it is difficult to organise these multi-faceted issues, not to mention address them with structural applicable approaches and appropriate answers.

Listing firsts and achievements in DH is no longer an option to solve questions about ‘what is DH’ and ‘who we are’. With the expansion and development of DH, it is practically impossible to go through various fields and enumerate related works and projects, and this agrees with what McCarty stated, ‘for computing to be of the humanities as well as in them, we must get beyond catalogues, chronologies, and heroic firsts to a genuine history’ (McCarty, 2008, p. 255).

This PhD study adopts a well-developed research model proposed by Zuccala – the invisible college – and systematically links large-scale data visualisations to detailed individual cases (Zuccala, 2006, p. 160). It employs both quantitative and qualitative methods and attempts to draw a bigger picture of DH not only at its current status but also over the historical periods of its development, not only through a macroscopic view of the whole field but also through microscopic vision of individual cases.

1.3.1 Invisible college

A discipline thrives when it has formed an academic community of scholars actively interacting and communicating ideas among their peers even though they might be geographically located at various affiliations across the world (Crane, 1972; Becher and Trowler, 2001a, pp. 41–42). Such communication is the ‘essence of science’ (Garvey, 1979), and such community is called an invisible college which is described as a connected system of a group of scholars within an academic field (De Solla Price and Beaver, 1966).

The phenomenon of ‘invisible college’ has been around for four-hundred years of history (Kronick, 2001, pp. 28–29), and it is more often discussed by scholars in the contemporary era (De Solla Price and Beaver, 1966; Crane, 1972; Gmür, 2003), although they do not seem to come to an exact agreement on a definition. An invisible college of any discipline is usually believed to be an organised system for scholars with some degree of predictable behaviour, such as communication, collaboration, sharing and exchanging ideas, and organising academic events (Crane, 1972, p. 33; Griffith and Mullins, 1972, p. 960). It normally grows and becomes larger when scholars communicate topics, interact with each other at certain events, such as conferences, and exchange news. Over time, the publication and interpersonal networks that were previously invisible to the wider academia might become more visible through authorship details, citations, acknowledgements, academic events (Zuccala, 2006, pp. 152–168), and more recently, through social media interactions (Ross et al., 2011, pp. 214–215). Thus, the concept of the invisible college offers a fertile land for disciplinary (and interdisciplinary) studies to build research models that could be used to systematically study academic fields.

Research modelling is not only important to existing fields such as the history of science or library and information studies, but also to new disciplines like DH, because it provides structural approaches to study and formulate a complex but comprehensive methodology. Some research models could even be able to reach the status of a theory through continuous development (Carrington et al., 2005, p. 3).

In 2006, Zuccala introduced a new theoretical framework to model the invisible college, and further refined it in 2009 (Zuccala, 2006, p. 155; Zuccala and van den Besselaar, 2009, p. 120). It is recognised as a ‘theoretically well-grounded framework’ and tested by other informatics studies (Teixeira, 2011, p. 2). In many research areas, this model has been employed to help scholars discover domain structures and scholarly communications. Examples include Entrepreneurship (Aldrich, 2012), Management and Organisational Studies (Vogel, 2012), Consumer Behaviour Research (Tu, 2011), and Organic Chemistry (Todres and Todres, 2009, p. 442).

Zuccala’s model provides a series of steps that consists of three sets of approaches, i.e., *subject specialty*, *social actors*, *information use environment (IUE)*. *Subject specialty* indicates a research discipline (or field) that reflects disciplinary rules and research problems, whereas *social actors* represents the communicating scholars connected via their social ties (information producing and sharing), and *information use environment (IUE)* represents the wider contexts and backgrounds of researchers, such as working institution and wider environment.

All three approaches are closely related and essential to each other (Lievrouw, 1990). *Subject specialty*, in her proposed model, is reflected in publications and citations. *Social actors* can be studied through interpersonal interactions, relationships and networks (i.e., collaborating scholars via different social channels). *IUE* aims to investigate scholars’ backgrounds and provides more complete contexts to explain deeper interconnections with the community formation. For example, scholars working within the same institution might have an impact on one another (Zuccala and van den Besselaar, 2009, p. 120).

Zuccala assigned each approach to a different set of methods and each combines both quantitative and qualitative analysis – bibliometric method (for *subject specialty*), sociometric method (for *social actors*), and ethnographic method (for *IUE*) (Zuccala,

2006, p. 156; Zuccala and van den Besselaar, 2009, pp. 117–120). Figure 1.3 shown below is the original graph in (Zuccala, 2006) that demonstrates the organisational structure of the invisible college research model highlighting the interrelationship among its three components.

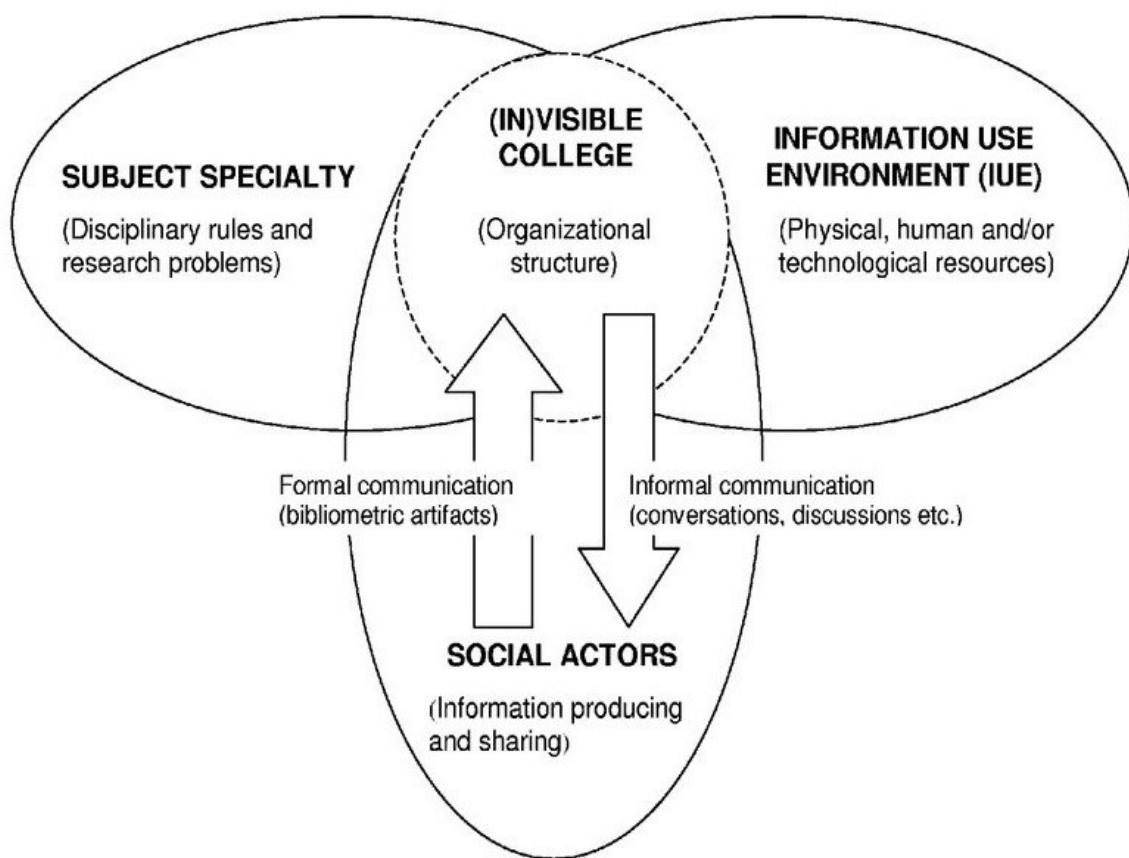


Figure 1.3: The original ‘invisible college’ research model (Zuccala, 2006) that demonstrates the organisational structure of the invisible college research model highlighting the interrelationship among its three components.

Bibliometric method typically focuses on academic publications to find out the formal ways of interaction (e.g., scholars’ citation behaviours, most influential authors according to the citations), while sociometric analysis helps to discover interpersonal relationships of communication (e.g., researchers’ personal conversations at certain events, or interpersonal interactions via online platforms) (Zuccala, 2006, pp. 156–166). Ethnographic (or environmental) method of analysis is more situated at detailed ways of interaction, and qualitative research methods are usually expected (e.g., immersive field studies, observational studies, interviews, storytelling, content analysis

of narratives). However, such ethnographic approaches focus more on small-scale cases that are difficult to conduct when a field expands globally, or the number of scholars exceed a certain amount, e.g., a thousand. Therefore, this study chooses quantitative methods to engage with environmental factors, and the detailed approaches will be explained in the next section (1.3.2 DH invisible college).

The three components in Zuccala's model (i.e., subject specialty, social actors, information use environment) match the three research questions that are shaped in section 1.2 (Research questions), i.e., *subject*, *scholar*, and *environment*. Nevertheless, it should be noted that there are differences between the two sets of concepts. Each component in the original model deals with all three research questions from its own perspectives (i.e., bibliometric, sociometric, ethnographic perspective). For example, through bibliometric analysis, one can explore all three questions. Bibliometric analysis can study the DH topic distribution (*subject*), the co-authorship (*scholar*) and backgrounds of these scholars (*environment*) from its formal communication channel. It is the same with sociometric and ethnographic analysis.

1.3.2 DH invisible college

To evaluate whether a discipline may function as one or more invisible colleges, Zuccala's research model provides five requirements (Zuccala, 2006, p. 157):

1. How young is the subject specialty? (It cannot be too old in the sense that many of the foundational scholars are either deceased or no longer publishing in the area).
2. Does the subject specialty fit within an identifiable indexing or classification system? (e.g., the American Mathematical Society Classification Code).
3. Is there a Web page associated with this specialty where participants have access to current research information (e.g., preprints), including information about national/international conferences or workshops?
4. How many scholars are identified with this specialty area (if there are too many scholars, the subject specialty is likely too large for members to know one another and interact informally)?
5. Are the scientists/scholars distributed worldwide? (There is no rule that invisible college members must be international; however, if they are and

there is evidence that they meet at selected conferences, there is an added richness to the kind of interpersonal communication that takes place.)

Studies have shown that DH has strong formal and informal networks and fulfils the above requirements (centerNet, 2018; Grandjean, 2016, para. 3), and thus, is an invisible college. DH has a history around 70 years (Jones, 2016, pp. 1–12; Nyhan and Flinn, 2016, p. 1). More libraries and librarians are supporting DH within their classification systems (Bryson et al., 2011, para. 1), and it is not only for DH publications in English, but also in other languages, such as Chinese (CNKI, 2019) and Spanish (Suárez, 2010). DH has a widespread research influence around the world (Terras, 2012a, para. 3), and scholars are interacting informally via social media more often than people from many other disciplines (Quan-Haase et al., 2015a, p. 3; Ross et al., 2011, p. 214).

An invisible college usually belongs to a discipline as a subgroup, while a discipline is not necessarily an invisible college (De Solla Price, 1963, pp. 12–14; Hagstrom, 1970, pp. 87–88; De Solla Price, 1986, pp. 5–6). However, given the difficulties to divide DH into distinct subgroups, and due to its very complex and highly debatable boundaries and significantly interdisciplinary scope, this study follows Quan-Haase's work (Quan-Haase et al., 2015b, p. 3) that sees the DH community (or communities) as one invisible college (i.e. the DH invisible college), even though there are many subfields existing within the 'big tent' of DH.

In general, the invisible college model is suitable to examine DH as a framework (Quan-Haase et al., 2015a). As Liu suggested, methodological frameworks borrowed from science studies have an important part to play in advancing the maturation of DH (Liu, 2013, p. 416), and previous efforts have agreed the feasibility of applying the invisible college research model to study DH and its community, e.g., (Bowman et al., 2013; Burton, 2015; Quan-Haase et al., 2015a).

However, some of the original approaches suggested by Zuccala are not as practical as they were first proposed, and the current study has refined the model and introduced new approaches. It has been 13 years since the model was introduced, and although approaches like co-citation analysis (for *subject specialty*) are still useful

and even gaining more attention, other approaches might need alternatives (e.g., conference co-attendance for *social actors*).

According to Zuccala's model, the pattern of conference attendance was designed as an approach to study social (i.e., informal) connections among scholars in an invisible college (Zuccala, 2006, pp. 159–161). Attending conferences is an important way to develop informal connections that provides unique benefits (Harrison, 2010, p. 263), but no assumptions can be made that two authors attending the same conference indeed had communication, especially as major DH conferences have become increasingly large during recent years. For instance, the DH2016 conference held at Krakow in Poland had 902 delegates from 45 different countries, and there were eleven sessions running concurrently throughout most of the conference⁴. This study, hence, will not include the pattern of conference co-attendance to examine the DH *social actors*.

Instead, a co-authorship network can help to explore direct scholarly collaboration. Although collaboration and co-authorship do not necessarily have a one-to-one or collocative relationship, co-authorship can be viewed as one important indicator of collaboration (Nyhan and Duke-Williams, 2014a, p. 387). Co-authorship approach was only partly involved in the original *social actors* section of Zuccala's model only as a verification method, but it will be added to the current study as the second part of bibliometric analysis to reveal more insights.

On the other hand, social media, and Twitter in particular, also plays an important part in the formation of the DH intellectual landscape and social community. When Zuccala proposed the model in 2006, there were not many social media platforms available for large-scale data analysis, and Twitter was not a useful data source to present an invisible college. In recent years, however, sociometric analysis (or sociometry) on social media has become one of the common methods to study online topics and communities. The DH community is believed to be among the early adopters of social media and online communications. Apart from the long-standing discussion forum Humanist (Nyhan, 2016), DH scholars have embraced various social media, such as

⁴ Please see the link of the DH2016 conference program for more information, available at: <https://dh2016.adho.org/schedule/>

blogs, Twitter, and Facebook, to communicate with their peers and collaborators. These online activities and academic interactions are informal communications that reflect and complement the formal research practices and publications. Therefore, by collecting data from Twitter to construct a hashtag co-occurrence network and a co-retweet network, this study also explores the DH *subject* and *social* structures from a social media perspective.

The third part of the model – *information use environment (IUE)* has also been refined as mentioned earlier. *IUE* was originally set to represent ‘physical working space’ such as scholars’ affiliation and co-working environment, although Zuccala also admitted it could be ‘fundamentally ambiguous’ (Zuccala, 2006, p. 164). Interviews and field work were proposed by Zuccala to investigate such scholarly working conditions. Nevertheless, these ethnographic approaches focus more on small-scale cases that are difficult to conduct when a field expands globally, or the number of scholars exceed a certain amount, e.g., a thousand. Instead, wider social elements (e.g., gender, race, language, country, diversities) might make a greater impact on scholarly information behaviour and communication patterns. For example, two scholars from the same institution might have similar characteristics, however, scholars that speak the same first-language, or are the same gender sometimes have stronger collaborative links (even internationally) elsewhere than with colleagues in the same department (Badar et al., 2013; Chen and Hsueh, 2013; Li et al., 2013, p. 1518; Tello, 2016; Mäkelä and Tolonen, 2018). These elements of scholarly background have attracted many debates and discussions that are central to DH, and it is necessary to address the correlations between these factors and different trends and behaviour. Although these diversity elements were not listed when Zuccala proposed the original IUE method, they are important social factors in a broader academic environment, and sometimes they have deeper and more profound influences on scholars than their physical working space. Therefore, this study takes the gender, affiliated country, and language of DH scholars into consideration for their popularity among recent DH debates and practicality of quantitative approach.

In conclusion, by employing the refined methodological framework, this study not only focuses on solving the research questions shaped in section 1.2, but also commits to act as a ‘tentative’ visual guide to the current practices and future trajectories in order

to provide a comprehensive view of DH scholarly production and social community via networks.

1.3.3 Chapter summary

By improving the invisible college research framework, this study organises its chapters as the following:

Chapter 1 (Introduction) has reviewed previous questions and fundamental issues in DH, and accordingly, it has formed the research questions and introduced the methodological framework of this PhD study.

Chapter 2 (Literature Review) will review and discuss the previous efforts devoted to the research questions (i.e., DH subject, scholar, environment) with emphasis on similar methods applied as proposed in the methodological framework in chapter 1.

Chapter 3 (Methodology) will present a detailed reasoning on approach selection and research design within each component of the methodological framework. It will extend the discussion about the refined invisible college model to a procedure construction of step-by-step guidance with some approaches that are currently available but have not yet been applied to DH.

Chapter 4 (DH Bibliometric Network Analysis) will demonstrate the construction process and visualisation results of two DH bibliometric networks – author co-citation analysis (ACA) network and co-authorship network. At the end of each network, there will be a discussion to interpret the network and analyse the results.

Chapter 5 (DH Twitter Network Analysis) will provide a detailed construction procedure of two Twitter networks (hashtag co-occurrence network and user co-retweet network) and the visualisation results. Similarly, at the end of each network, there will be a section to analyse the results and present interpretation.

Chapter 6 (Conclusion) will summarise the interpretations and results and answer the three research questions proposed in chapter 1 (Introduction).

Chapter 7 (Reflection and future study) will review the whole study and discuss its importance, limitations and future study.

To the best of the candidate's knowledge, this study is the first to apply a well-grounded research model to systematically combine and visualise the DH knowledge and social structures. This research stands apart from existing DH debates, narratives, and disciplinary reviews as it conducts inclusive and representative empirical research among other bibliometric and sociometric studies. The resulting outputs will make a valuable contribution to the current debates about DH knowledge structure and the wider scholarly networks, and also to the ongoing discussions about the formation of scholarly communities. It has constructed timely and useful datasets that can be reused for other DH studies which will be published freely and openly in the UCL institutional repository as part of the Open Science Agenda, and the methodology applied in this study is compatible with studies of other evolving fields and disciplines.

2 Literature Review

This chapter reviews the broader literature within which to situate the three research questions shaped in chapter 1. The review and discussions are made under the structure of the invisible college and with emphasis on the related quantitative approaches proposed in the refined methodological framework (as in section 1.3.2) from three aspects – DH subject specialty, DH social actors, and DH information use environment (IUE).

With the help of qualitative analysis, quantitative approaches not only provide a ‘vision of the whole’ but also bring together considerable detail of individual elements that were once too many and too complicated to specify (Börner, 2011). In recently years, there have been efforts to try to quantify DH as a field and they have attracted increasing attention and interest. Some of them focused on analysing different sources (e.g., DH journals, conferences, grants, social media, curricula), some concentrated on different methods (e.g., statistical approaches, data visualisation, network analysis, topic modelling), and others tried to cover multiple aspects. An overview of some of the scholarship has been compiled by Weingart in his ongoing blog since 2016 (Weingart, 2018). Earhart also outlined some related scholarship in a global context that used quantitative methods (Earhart, 2018).

This chapter not only extends their endeavours by providing a systematic literature review from three perspectives, but also discusses improvements upon these studies and leads to the empirical experiments in later chapters. Section 2.1 (DH subject specialty) reviews the previous quantitative studies that addressed questions about the DH subject specialty. Section 2.2 (DH social actors) discusses the existing scholarship on the DH social actors research question (e.g., collaboration behaviours and publishing patterns). Section 2.3 (DH information use environment) comments on the previous studies that have researched on the DH IUE question.

The three sections are constructed according to the order of the three research questions; they are, of course, not exclusive, but overlap, and each section is presented with its own focus. Some studies reviewed in this chapter cover two or all three aspects; and are discussed in each relevant section but from a different perspective. For instance, the ADHO conference serial studies conducted by Weingart

addressed them all, i.e., the DH conference topic (subject specialty), the co-authorship of the participants (social actors), and the gender and affiliated country distribution (information use environment) (Weingart and Eichmann-Kalwara, 2017).

2.1 DH subject specialty

As mentioned above, different scholars understand DH differently. Yet, even empirical studies may present quite different landscapes of DH. Depending on the data scale, the method, and the type of sources that different studies use for their data collection, the results and representations of the DH subject specialty vary significantly.

This section reviews and discusses existing efforts to address the question of 'what research topics is the DH subject composed of?' via different types of data sources (i.e., journal – 2.1.1, conference – 2.1.2, Twitter – 2.1.3, blog – 2.1.4, and others – 2.1.5). Sections are ordered by their level of formality (where journal publication is the most formal source while social media is less formal⁵) (Abu Sheikha and Inkpen, 2010). Each of them presents one or more DH intellectual structure(s) that are not identical but complementary and comparable.

2.1.1 Journal⁶

DH has its own journals, and the first one, *Computers and the Humanities (CHum)*, was started as early as 1966. Other influential journals include, but are not limited to, *Digital Scholarship in the Humanities, DSH* (formerly known as *Literary and Linguistic Computing, LLC*) and *Digital Humanities Quarterly (DHQ)*⁷. The main issues of these journals are published in English, but there are also DH journals of other languages. For example, *Digital Studies / Le champ numérique* particularly encourages global,

⁵ The level of formality is decided based on the research conducted by Abu Sheikha and Inkpen in their paper *Automatic classification of documents by formality* (Abu Sheikha and Inkpen, 2010).

⁶ Before reviewing the bibliometric studies, it should be noted that certain differences exist between the term 'reference', 'bibliography', and 'citation' (Hellqvist, 2009, p. 310). However, this PhD thesis uses them interchangeably to refer to a list that links together sources, notes, and credits of materials presented in a publication. Nevertheless, wherever possible, this study will still differentiate the use of these three terms to the use of 'footnote' and 'endnote' that are very different in the contexts of both the humanities and sciences.

⁷ The former was founded in 1986, whereas the latter was launched in 2007 and is fully open access.

multi-cultural, and multi-lingual submissions, *Le foucaldien* journal is published in multi-languages, *Zeitschrift für digitale Geisteswissenschaften* journal is for DH articles in German, and there is also a DH journal in Chinese (i.e., 數位典藏與數位人文, *Journal of Digital Archives and Digital Humanities*).

Apart from journals, a great number of books and edited volumes have been published under the 'DH' label to assist research, practice, teaching, library management and founding infrastructures, etc., especially from 2005, e.g., (Gardiner and Musto, 2015; Gold, 2012; Gold and Klein, 2016; Schreibman et al., 2016, 2004a; Terras et al., 2013).

The increasing growth of DH publications provides an opportunity to review the field by examining its productions. Journals are often more favoured by empirical studies as data source rather than books or monographs. As far as the candidate is aware, all the existing studies on DH formal publication (although not many) used journals rather than books as the data source. This might be because journal articles have more timely content, structured bibliometric metadata, and a wider range of topics. Even though both represent the products of formal academic communications, the data collected from journals is often more efficient for an empirical approach. The bibliometric components of a journal article (e.g., author, title, keyword, or bibliography) form a connected system that could define the knowledge structure of a discipline and the academic formal communications between scholars in that discipline or research area (Tang et al., 2017, p. 986). The conference abstract is another type of publication that is less formal than the journal article but more formal than social media content, and it will be discussed in the next section (2.1.2)

At the time of writing, only a handful of studies have dedicated their attention to explore the DH subject and knowledge structure by analysing DH publications. In 2009, Wang and Inaba conducted the correspondence and co-word analysis on articles published between 2005 and 2009 in *LLC*, *DHQ*, and *ADHO* proceedings (Wang and Inaba, 2009a). They visualised the intellectual structure of DH by extracting 1,219 title terms from 548 articles and calculating the co-occurrence of the title words. They then generated the networks of the most frequent 82 words for the time range (2005 - 2009) as well as in each individual year (Wang and Inaba, 2009b).

According to their research outputs, by the time they presented this work (in 2009), the disciplinary nomenclature was moving from ‘Humanities Computing’ to ‘Digital Humanities’ (shown in Figure 2.1).

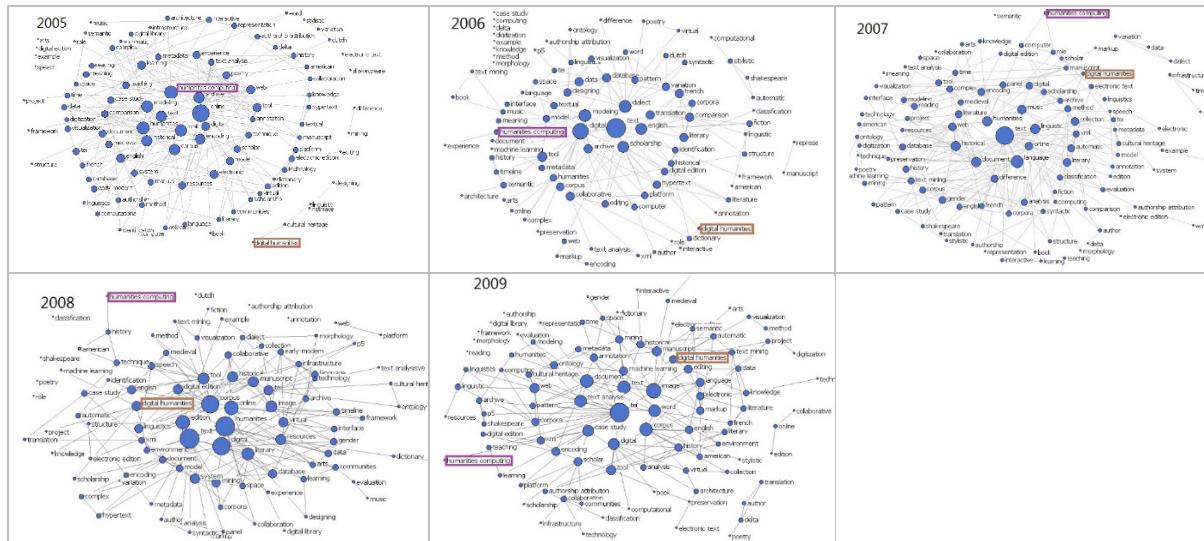


Figure 2.1: The co-occurrence network of high frequency words in 2005 – 2009 by (Wang and Inaba, 2009b, pp. 18–22), and node size represents degree centrality. Use in this thesis has been permitted by the authors⁸.

As shown on the networks generated by Wang and Inaba, the term ‘Humanities Computing’ (marked in purple) became less mentioned in article titles and was gradually replaced by the term ‘Digital Humanities’ (marked in brown) over the five-year period. These networks reflected a significant change in the field’s nomenclature in just five years, and provided quantitative evidence of this rapid change that affected not only the topics and scope of DH but also the values and identities of the community (Svensson, 2009).

Wang and Inaba also found, surprisingly, that there were no clear subfields in DH. This finding was unexpected because DH is usually described as interdisciplinary, collaborative, and with a broad range of topics under its ‘big tent’ (Hockey, 2004; McCarty, 2003b; Warwick et al., 2012). A high level of interdisciplinarity should, hypothetically, provide great opportunities for DH subjects to develop, group with related topics, and form subfields (Klein, 2015). Wang and Inaba did not give an

⁸ For high resolution image, please see Appendix G.

explanation for this surprising result but pointed out that an expansion of the dataset and controlled study of other fields were needed. Although they made the first step to visualise networks of DH intellectual structure, their data only covered the period between 2005 to 2009 (ending around 10 years ago). Putting it into the current circumstances, such data seems limited, and the time gap might explain why the network they generated was vastly different to what we currently perceive.

The lack of the latest and sufficient data seems to be the problem with other DH subject studies, too, although this is only because these studies were conducted relatively early. From around 2010, a group of scholars including Salah and Leydesdorff presented their series of studies of the DH knowledge structure based on journal data (Leydesdorff and Salah, 2010; Salah et al., 2010, 2010, 2015), and some of their early works also indicated unexpected results that are akin to Wang and Inaba's – intellectual structures of limited topics and narrow disciplinary involvement.

Firstly, in January 2010, Leydesdorff and Salah used the Arts & Humanities Citation Index (A&HCI) to evaluate the citation patterns of journals. In order to demonstrate that a journal was not the only unit that organised the intellectual data of a discipline, they studied DH publications by searching DH keywords (Leydesdorff and Salah, 2010, p. 18). They collected DH publications from Web of Science (46 documents) by searching titles that contained either 'digital humanities' or 'humanities computing' in 2009 (Leydesdorff and Salah, 2010, pp. 18–19). Although 46 documents was a small number of publications, and the time period (1975 – 2009) was relatively short compared to studies of other domains (e.g., Eom, 2003), they gathered 829 cited references that were from 81 different journals. By constructing a journal co-citation network, they found that DH was not as interdisciplinary as it was described. This finding agrees with Wang and Inaba that the subjects were not diverse enough to form clusters, and it showed that DH topics were only from a restricted number of fields, such as library and information science (34.6%), the application of computers in linguistics (10.9%), and computers and literature (6.1%).

Five years later, in 2015, Salah expanded the data range collecting 390 articles by keywords 'digital humanities', 'humanities computing', 'e-humanities' and 'computational humanities', and constructed the network of journals as shown in Figure 2.2 (Salah et al., 2015). It seems that with the growth of data range and time

period (i.e., five more years), more subjects and topics in DH are starting to show in the results.

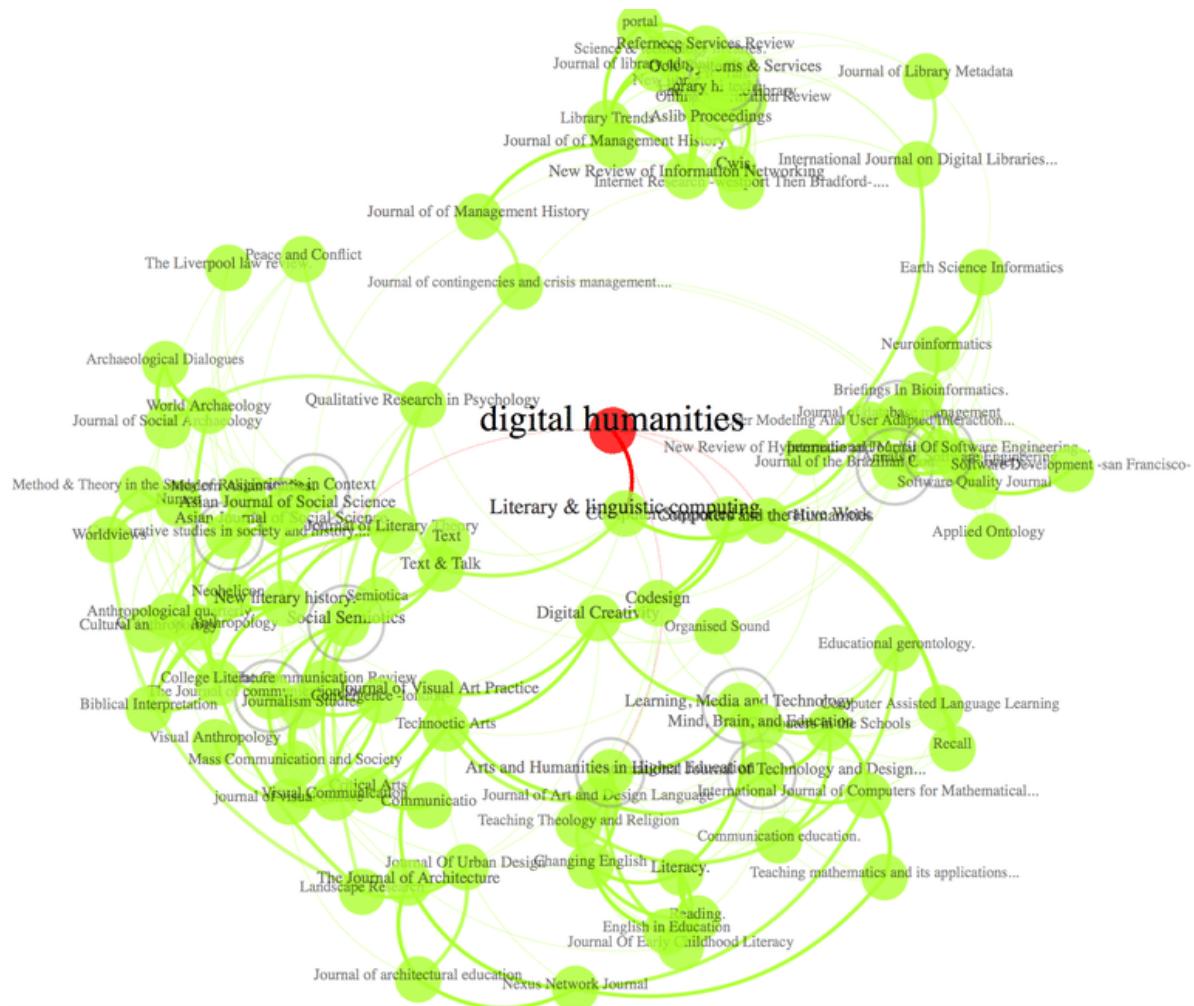


Figure 2.2: Contextual exploration of Digital Humanities (Salah et al., 2015, p. 83).

Figure 2.2 above shows a more complete picture about the position of DH in a wider scholarly environment and the relationships of DH with journals in other fields. At the top of the network, there was a cluster of library journals, while on the right hand-side there were journals in computer and information science. The largest cluster is on the bottom-left with journals from various humanities areas (e.g., media studies, literary history, arts and humanities).

This series of works led by Salah and Leydesdorff demonstrated that the representations of DH subjects can be changed and developed with the growth of data scale and coverage. Yet, can simply increasing the data range present a more comprehensive structure of DH subject? What amount of data is considered 'enough'?

Tang and his co-authors conducted a large-scale DH bibliometric analysis (Tang et al., 2015, 2017). By searching keywords on Scopus and retrieving articles published in six DH journals, they collected 2,509 publications (including articles, conference papers, books, book chapters) from 1989 to 2014 and constructed three networks (Tang et al., 2017, p. 990). Through these works, they demonstrated that more subjects and disciplines could be found from DH related publications as the data scale increased and more methods were introduced, especially during recent periods. However, they also found that DH was still a 'small world' where everyone knew everyone.⁹ These two findings seem contradictory with DH subjects being more global and diverse while at the same time, being in a dense and small world where every subject is closely connected.

More specifically, they visualised an author bibliographic coupling network (Figure 2.3) that demonstrated a very diverse group of subjects from a mix of dynamic disciplines. Various sub-fields and research topics were identified across computing and humanities disciplines. For example, by examining Figure 2.3, they found that DH's degree of knowledge diversity was higher than other humanities disciplines, and sub-fields grew more apparent. Tang argued that it demonstrated a high possibility that DH had generally developed as a diverse but cohesive discipline with high global reach.

⁹ A small-world network is a common type of mathematical graph in which most nodes are not only neighbours of one another, but also neighbours of most of the nodes that can be reached in the network (Watts and Strogatz, 1998). One of the small-world network properties is known to be power-law obeying degree distribution (Bork et al., 2004).

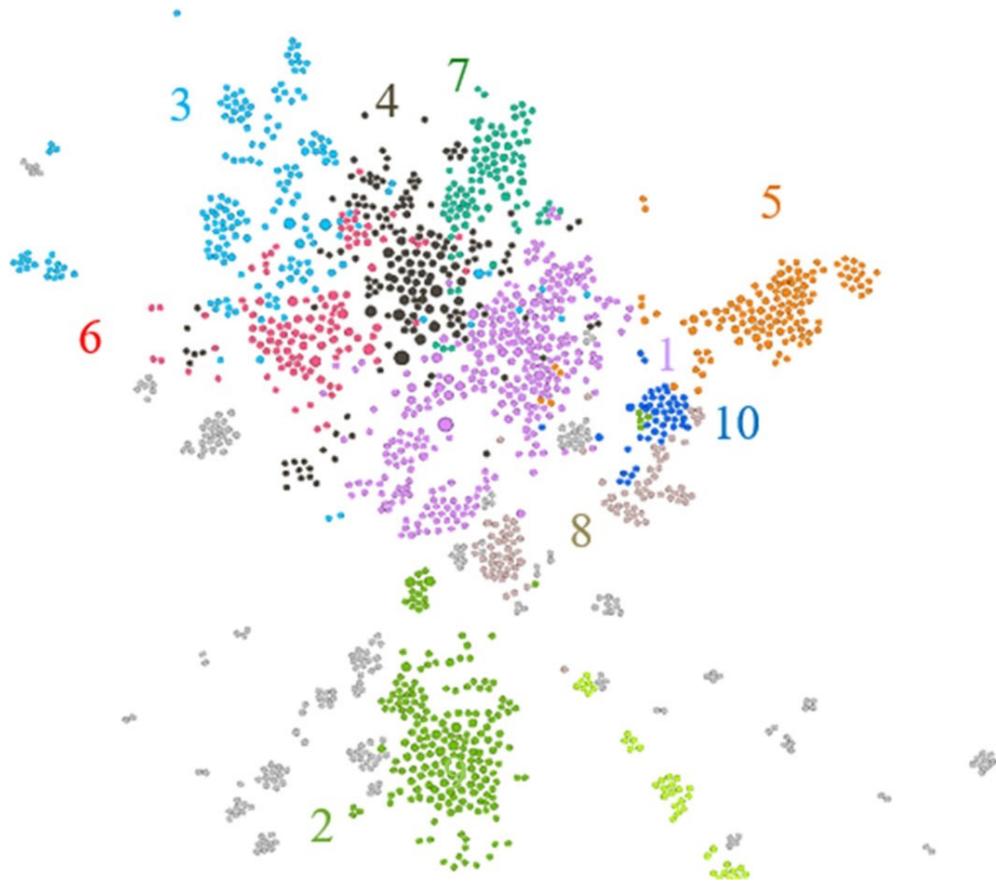


Figure 2.3: Visualisation of modularity classes in author bibliographic coupling network (Tang et al., 2017, p. 1001), and colours indicate different clusters that match the numbers of the same colour. Use in this thesis has been permitted by the authors.

Nevertheless, another network (document co-citation analysis) generated by the same authors found the opposite. Figure 2.4 below shows that there are very dense local subjects at the centre of the DH network. Most subjects and topics were grouped together without clear separation, and they believed that this was a typical sign of the ‘small world’ model.

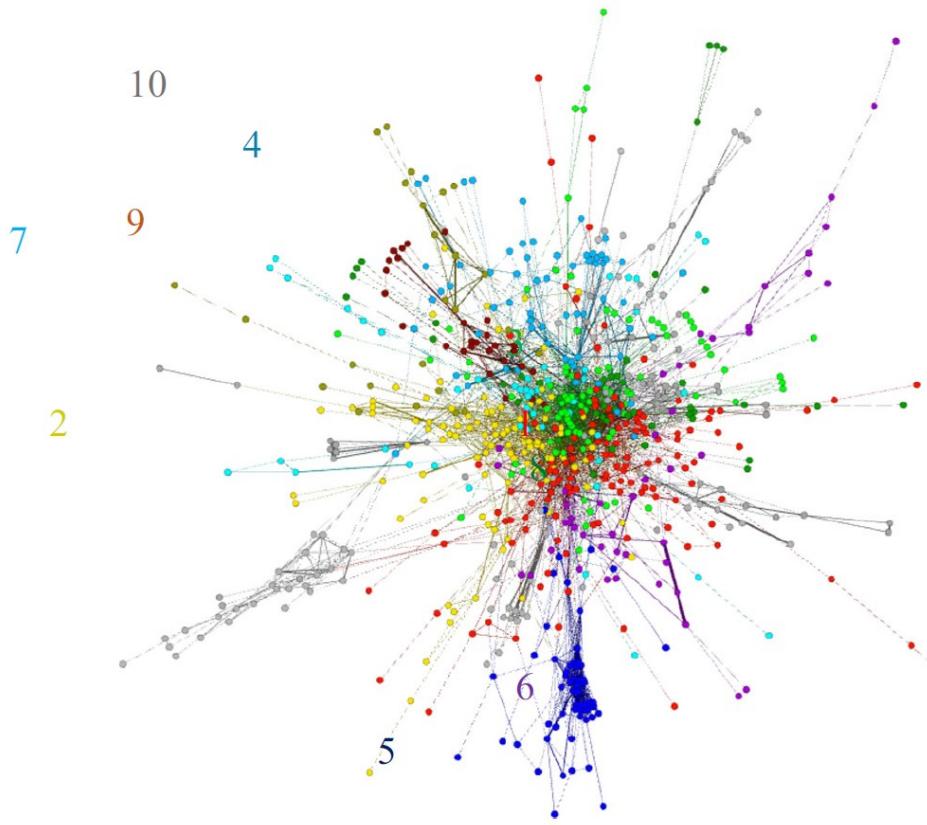


Figure 2.4: Visualisation of modularity classes in document co-citation network by (Tang et al., 2017, p. 988), and colours indicate different clusters that match the numbers of the same colour. Use in this thesis has been permitted by the authors.

How could DH subjects become more and more diverse and interdisciplinary while at the same time form a small and highly connected world? Why did the same dataset analysed by two different methods present two contradictory results? As Tang pointed out, although based on the same data, different network methods could produce distinct topologies, and caution needed to be taken. There are many kinds of network approaches, and each with different emphasis and calculation. Some might focus on the knowledge base of a field, while others could focus on its recent intellectual structure.

As Tang noted, their lack of domain knowledge at the micro level (e.g., DH scholars, their backgrounds) was a limitation of their work. Without a comprehensive understanding of DH, its history and development, it is difficult to interpret the visualised networks, empirical results, and make further contributions to the disciplinary debates. Tang also indicated the lack of a comprehensive bibliographical

dataset of DH publications (Tang et al., 2017, p. 1003), an issue that this current thesis addresses (in chapter 3 Methodology and chapter 4 DH Bibliometric Network Analysis).

2.1.2 Conference

Compared to journal publications, conference abstracts are timelier and less formal, but the acceptance rate is competitive (Weingart, 2016a). DH societies across the world (e.g., ADHO and its member organisations) hold various conferences.¹⁰ However, the studies that will be discussed below show that only a handful of topics dominate those conferences, and depending on the conference location, language, and data source, the popular topics can be very diverse and dynamic or very concentrated and unitary.

In 2006, Terras made the first attempt to quantitatively analyse the field's international conferences when the field was still mostly referred to as 'Humanities Computing' (Terras, 2006). By analysing the word frequencies in the abstracts of the annual ACH/ALLC conference¹¹ (1996 – 2005, excepting 2003), she identified the most popular topics – computational analysis of text with the emphasis on the language, words, and documents (Terras, 2006, p. 236). With further analysis on around 250 presenters at ACH/ALLC 2005, she found that they were from a limited number of disciplines. Apart from specific DH (or Humanities Computing) centres, these presenters were mainly from library and information studies, English, and linguistics disciplines.

Six years later, Terras presented another quantitative work (Terras, 2012a). In this work, she demonstrated the geographic distributions of DH centres across the world, the different disciplinary indicators (e.g., numbers of journal subscriptions, followers on Twitter, access statistics to DH resources, number of attendances at DH events, etc.), and information about their resources, such as investment and funding. Although

¹⁰ ADHO, for example, and its annual conference provides an international platform for DH scholars and unite dynamic networks of scholars regardless of their home disciplines and research topics (Siemens, 2016, p. xxv).

¹¹ The conference was called the annual Association of Computing in the Humanities and the Association of Literary and Linguistic Computing Joint International Conference (ACH/ALLC) before 2006, and the conference later became the international DH conference held by ADHO (the amalgamation of ACH and ALLC).

these numbers witnessed the increasing growth of DH, there was no clear evidence of increase in subject diversity or topic growth.

Starting in 2012, Weingart collected data from the ADHO annual conference and published a series of blogs and publications on the conference analysis that covered topics, scholars, geographic distributions, and peer-review bias of the conference submissions and programmes from 2012 to 2016 (Weingart, 2012, 2013a, 2013b, 2014a, 2014b, 2014c, 2015a, 2015b, 2015c, 2015d, 2015e, 2016b, 2016a; Weingart and Eichmann-Kalwara, 2017).¹²

Weingart's analysis indicated that the DH subject developed from project-based to principle- and skill-focused topics, whereas the popularity of the most prominent topics remained the same, such as literary studies (20%), text analysis/mining (20%), archives (19%), and visualisation (18%). This finding agreed with the results produced by Terras. Although project-based subjects declined (e.g., interface and user-experience design, scholarly editing), the main focus of DH subject was still text (e.g., literary studies, text analysis).

The DH conference as a data source has its own limitations. Its subject distribution is relatively unstable due to the annual change of the conference location. After examining the conference abstracts of DH2015, Sydney, Australia, Weingart found that not only was the author country distribution more diversified, the topic distribution was affected too. For example, the proportion of Asian studies was nearly doubled compared to the previous year (Weingart, 2014a). This change agrees with many related studies on scholars' conference attendance motivation and decision-making that location is an important factor that considerably affects conference activity (Oppermann and Chon, 1997; Severt et al., 2007; Mair and Thompson, 2009); such distinction may be more obvious at regional DH conferences.

Regional conferences play a significant part in forming the DH global perspective, even though many of them are not associated with ADHO (Galina, 2013). Studies show that while the DH subjects of regional conferences could be concentrated on just

¹² This section will focus on Weingart's contribution to the DH subject research while his study on co-authorship and community diversity will be discussed in section 2.2 (DH social actors) and 2.3 (DH information use environment).

a few topics or very diverse across hundreds of topics, the distribution was associated with the conference location and language.

Chen and Hsueh analysed papers from the largest annual Sinophone DH conference, the International Conference of Digital Archives and Digital Humanities (DADH), and found that most topics were from humanities disciplines, and 64.6% of the papers focused on subjects dating back to the 17th century (Chen and Hsueh, 2013). Most papers studied the topics that were related to Taiwan (49.5%) or mainland China (21.36%), and a large proportion of them were history studies (31.18%) of historical materials, events, bureaucracies, and Chinese people in the Qing Dynasty (1636-1912). Other topics included culture (10.8%), politics (11.58%), literature (9.59%), geography (7.81%), linguistics (7.1%).

Among all the Sinophone regions, Taiwan holds the largest collection of Chinese historical objects and artworks (Fong and Watt, 1996). Many collaborative digital projects (e.g., digitisation, cataloguing) across different Sinophone regions have been conducted to make their collections more accessible (Wang, 2016). Therefore, it is not surprising to find that the topics at DADH are predominantly about Chinese historical studies.

However, not all the regional DH conferences have such a concentrated distribution of topics. By analysing the subject of DHd2018 (the German Digital Humanities Conference), Henny-Krahmer and Sahle found a very diversified distribution of topics (Henny-Krahmer and Sahle, 2018). In their analysis, 763 keywords were assigned to the papers but 578 were unique. This means that the topics at DHd2018 were very diverse, with the highest proportion accounting for only 6.4% (annotation), followed by TEI (5.3%), digitisation (4.3%), digital edition (3.2%), ontology (3.2%), and one of the ADHO's popular topics – visualisation – only accounted for 2.7% of the papers. Even though such a great difference might be partly because there was no controlled list of keywords to be selected, the significantly diverse distribution within the German speaking DH community is notable.

A similar keyword-situation prevails in the Dutch-speaking conference context according to Kemman's analysis of submissions to the DHBenelux conference¹³ (Kemman, 2016a, 2016b, 2017, 2018). Kemman found that among more than a thousand unique keywords, 'digital' and 'data' remained the most used keywords while 'Dutch', 'new', and 'different' were 'fairly common', and it perhaps reflects on the fact that Dutch-speaking scholars might wish to have something different and new compared to the mainstream DH topics.

Mäkelä and Tolonen's analysis of the DH conference in Nordic countries (DHN2016 – DHN2018) suggested that the terms of the conference call for papers can potentially influence the subject-spread of the conference (Mäkelä and Tolonen, 2018). Topics specified in the conference call were: History, Cultural Heritage, Games and Future. Compared to the ADHO conference focusing more on literary studies and text analysis, the popular subjects at DHN were cultural-heritage collections, digital resources, history, linguistics, and GLAM (Galleries, Libraries, Archives, & Museums). Thus, the number of submissions on historical and cultural studies was three times higher than the number on literary studies.

Although most regional conference analyses were conducted using descriptive statistics to study the question of topics, they contributed greatly to the understanding of the DH global knowledge structure. The reviewed studies not only presented different local qualities of the DH knowledge, but also helped to connect local subjects to a wider international context. However, whether the conference location and theme have direct or correlated impact on the conference subject is still unknown.

2.1.3 Twitter

Although the DH community is very active on Twitter, few studies have explored its distribution of topics (apart from Moravec, 2018).

Aiming to study the exceptionalism in DH, Moravec collected tweets that included the hashtag #digitalhumanities on Twitter from March 2009 till the end of 2012 (Moravec, 2018). Combined with data analysis on various DH manifestos, she calculated the

¹³ It is an annual DH conference held by predominately Dutch-speaking countries in turn (i.e. the Netherlands, Belgium, and Luxembourg).

frequency of ‘bursting’ words, bigrams, trigrams, and other hashtags that co-occurred with #digitalhumanities, and found that the main subject that DH scholars were discussing on Twitter was about the central values of the field. For example, two most popular topics were ‘participating’ (with words such as ‘collaboration’ and ‘diversity’) and ‘helping’ (with words such as ‘encourage’ and ‘support’).

In her study, Moravec noted that Twitter played a significant part in the formation of the DH community. Based on her statistics, Twitter developed from merely a ‘networking platform’ in 2009, to a ‘platform for scholarly communication’ in 2010, and a ‘conversational platform between individual participants’ from 2012 onwards. By calculating the hashtags that co-occurred with #digitalhumanities, she found that museums were important contributors to the early community formation. This finding agrees with what Brennan argued in her blog titled *DH Centered in Museums?* that ‘no one imagined DH – as a constructed field of practice – centered elsewhere’ (Brennan, 2015).

In general, the DH subjects discussed on Twitter were mostly ‘forward-looking’. As Moravec mentioned, ‘digital humanities is both “a burgeoning community” but also a “social utopia”, and ‘it seems we may have proselytized too hard and or have been heard as promising too much’ (Moravec, 2018, pp. 189–191). She, therefore, urged users to stop talking about DH in utopian terms (e.g., ‘revolutionary’ or ‘new’) and provided a modest manifesto to go back to its central values.

As reviewed, the topics discussed most frequently in journals and conferences are about research materials (e.g., text, data, literature), methods (e.g., data mining, visualisation), and projects (e.g., digitisation, libraries). Although discussions on values did start to emerge in recent years (e.g., diversity as a topic keyword appeared at the ADHO conference in 2016), they accounted for a small proportion and varied based on different datasets. Moravec’s findings on Twitter are very different from the results derived from journal and conference analysis. This may be because Twitter specifically provides an immediate conversational platform that facilitates interactive and timely information exchange (Ross, 2012), in contrast to the relatively lengthy process of journal and conference submission. Topics such as ‘diversity’ and ‘collaboration’ often attract various views and arguments that need continuous debates, although they are in journals, too. Some studies indicated a strong requirement for

such values (e.g., 'diversity') and there are still conflicts between DH practices and 'structural misogyny and racism' that need to be addressed (Losh and Wernimont, 2019, p. ix). These topics about values can be communicated in a more timely way on Twitter with ongoing conversational threads.

Secondly, the distinct subject matter on Twitter might be due to the data collection. In Moravec's study, only tweets that included the hashtag #digitalhumanities were collected. This means that all the DH tweets that did not have #digitalhumanities are excluded in the analysis, and there are many tweets about DH that have not included such a hashtag. Instead, they might use #DH, or other languages (e.g., #humanitésnumériques), or other DH event and project hashtags. In addition, tweets about DH research and projects might like to use #textanalysis, #dataVis, or #digitization rather than just #digitalhumanities. Yet, this assumption is made based on individual examples. A comprehensive examination of DH tweets that includes more hashtags is needed to investigate this matter.

Moreover, this subject difference could also be because the users on Twitter are not the same group of scholars that publish in journals and conferences, although there might be considerable overlap. For example, Holmberg and Thelwall found that it was difficult to identify the highly cited DH scholars on Twitter, and this was partly because many of them did not use Twitter or were already deceased (Holmberg and Thelwall, 2014). Such a gap also reveals some scholarly resistance and critiques towards social media that can be further studied (Sugimoto et al., 2017, p. 2038).

2.1.4 Blog

The blog is also a home to DH scholarly communications, and DH was even described as the 'blogging humanities' by McPherson in 2009 among all the other social media (McPherson, 2009). The DH blogosphere is very closely related to the daily interactions of the DH community and is an important part of the DH intellectual realm. The blog communications overlap with both formal and informal channels, and this makes blogging an interesting scholarly activity that gradually changes the nature of scholarly publishing and communication (McPherson, 2009, pp. 119–121).

In 2011, Meeks firstly explored the subject of DH on blogs along with other documents (Meeks, 2011). By web searching and private requests, he obtained around 50

documents and rendered a co-occurrence network. Although he identified around 20 topics based on word frequencies, there was no strong connections among them. This might be because the dataset is relatively small and, thus, not able to extract significant links among the topics.

In 2015, Burton expanded the dataset scale and collected 106,804 individual posts across 396 DH blog sites (1995 – 2013). By applying the theory of informal scholarly communications coined by Menzel – ‘trace ethnography’ – with topic modelling methods (Menzel, 1968), he demonstrated the important role of blogging in the formation of the DH intellectual landscape.

According to his results, popular DH-related topics on blogs can be classified into four categories. Apart from ‘junk’ topics and ‘non-English’, the most popular topic (32%) is ‘quasi-academic’ that included formal and mainstream humanities research and subject matter. The second (20%) is ‘para-academic’ (i.e., ‘research questions that were unique to DH and had otherwise no place in other disciplines’), while the third (14%) is ‘meta-academic’ (i.e., ‘disciplinary discussion and administration of DH’). Lastly, 10% of the blog subjects were about ‘enabling scholars to carry on certain studies on blog’.

Burton’s results showed that the blog includes both formal and informal academic themes, and he also found that while people were talking about academic themes both formally and informally, posts that adopted formal academic discourse often had more popularity than ones with informal contents (Burton, 2015, p. 143). Blog readers preferred formal structured posts, and this is not surprising as most audiences were academics.

What about blogs with more authority and specifically topic oriented? The Humanities, Arts, Science, and Technology Alliance and Collaboratory (HASTAC) and Hypotheses services offer blogging platforms with authority derived from their organisations.¹⁴

¹⁴ HASTAC is an online community that also enables social networking among scholars. It was founded in 2002 and by using free content management system, users could express their ideas about DH, media and communication. Hypotheses is a blog platform that enables the academic publishing and communication between two French institutions and universities, although contributors to it are not necessarily linked to these institutions and universities; founded in 2004, it has a clear focus on mainstream humanities, such as history.

Puschmann and Bastos collected 14,046 English-language posts (July 2006 – June 2012) from HASTAC (7,269) and Hypotheses (6,777), and they employed co-word analysis and topic modelling to discover the keywords that were related to DH works there (Puschmann and Bastos, 2015).

By constructing a term co-occurrence matrix, they revealed a preference for using humanities-related terms on blogs instead of ‘digital’ terms, and they identified different topics but mainly related to tones discovered from formal publications, i.e., Burton’s ‘quasi-academic’ category, ‘topics whose subject matter touched upon themes resembling formally published scholarly communication’ (Burton, 2015, p. 141). For example, as shown in Figure 2.5, the network of HASTAC illustrated a clear separation between different topics. The topic distribution of this network is similar to the network constructed by Salah et al., based on 390 journal articles in Figure 2.2 (Salah et al., 2015, p. 83), section 2.1.1. Topics such as ‘archive’, ‘literacy’, ‘pedagogy’, ‘library’, and ‘digital media’ account for the greatest area in both networks.

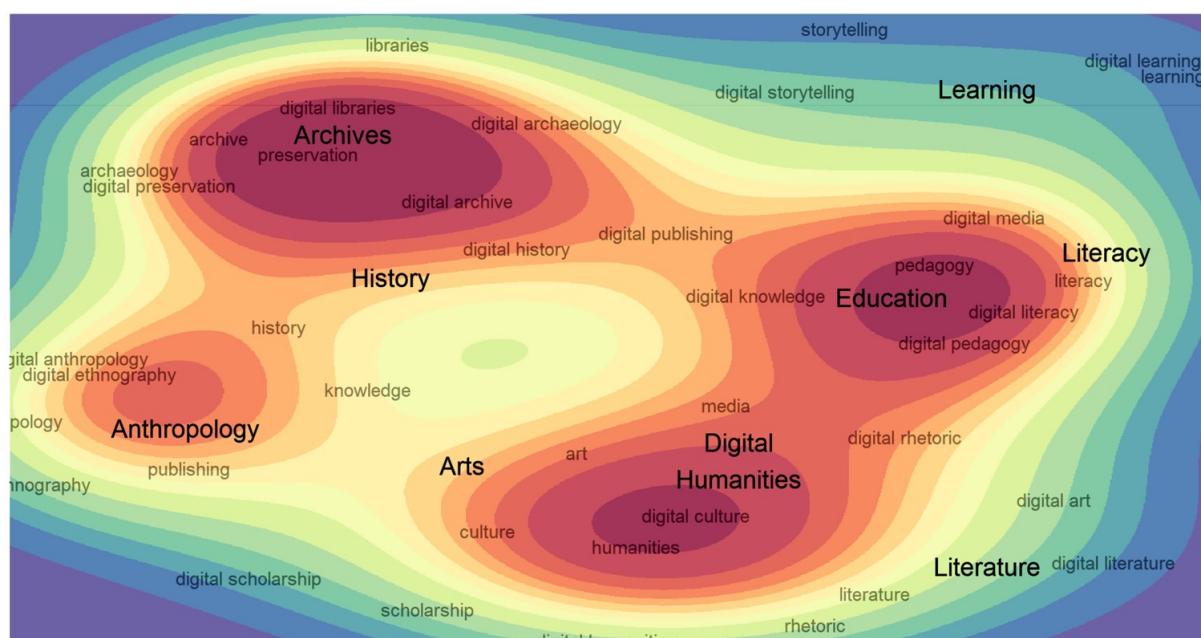


Figure 2.5: HASTAC density map of co-occurrence network on humanities-related terms (Puschmann and Bastos, 2015). The image used in this thesis is under Open Access CC 4.0 licence.

Additionally, they found that the topic distribution of HASTAC and Hypotheses are different. HASTAC has four main topic clusters that are separated while there was

only one cluster based on topics extracted from Hypotheses, and this was due to different regional and language focuses.

When we compare Burton's blog study to Puschmann and Bastos', we find that although these two blog studies are published in the same year and with the same topic modelling focus, the results are different. Burton's results showed a diverse range of topics while the networks of Puschmann and Bastos contained similar research topics to journal analysis. The distinction in their data sources might have contributed to such difference. Burton collected his blog data from *Digital Humanities Now*, while Puschmann and Bastos focused on HASTAC and Hypotheses. Thus, it seems that different blog datasets result in different subject structures, and it appears that people blog about different things on different platforms.

In general, the blog has its unique advantage of combining both formal topics (e.g., journal DH subject) and informal topics (e.g., Twitter DH subject). Nevertheless, depending on what type of blogs we harvest the data from, the topics on blogs can vary significantly from a very diverse range that covers research, education, administration, and informal ongoing discussions, to very focused research-oriented topics that are similar to themes found in formal publications. This variation also applies to DH subject studies on other data sources, such as DH curriculum and discussion forum.

2.1.5 Others

Apart from the subject studies mentioned above, DH research and practice can also be found from other media, such as syllabi (e.g., curriculum at universities) and discussion mailing lists (e.g., the Humanist, TEI).

In 2011, Spiro conducted a small-scale analysis based on 134 DH syllabi from 2005 to 2011 and investigated disciplinary distributions, technical skills, and requirement patterns across different courses, assignments, and readings (Spiro, 2011). The results showed that English departments offered the most DH courses among all the other disciplines (27.6%), with History departments ranking the second (16.4%) and Media Studies as the third (15.7%). The DH centres only held 12% of DH-related courses, and Library and Information Science that was believed to be closely related to DH only held 5.2%. There were no DH courses found in Linguistics or Classics

departments. This subject distribution is different from the ones reviewed in previous studies.

Later, some scholars tried to address the question from another aspect using data from discussion fora. Rockwell and Sinclair calculated the relative frequencies of 'digital humanities', 'humanities computing' and 'computing in the humanities' in the *Humanist* (1987 – 2008) (Rockwell and Sinclair, 2016).¹⁵ They produced similar results to Wang and Inaba's finding on the development of the field's nomenclature (Wang and Inaba, 2009a), and found that the use of 'digital humanities' brought changes to the field. Three different periods were uncovered: 1987 – 1995 (humanities computing), 1996 – 2000 (transitional period), 2001 – 2008 (a shift to digital web services and collaborative projects). Although this study had results that were similar to those found in previous journal studies, they also discovered many informal topics that can be classified as Burton's 'meta-academic' label (i.e., 'topics whose subject matter is focused on maintenance and organization of a social group'), such as topics related to 'services'.

In 2014, McClure continued the analysis of *Humanist* and collected 27 years (1987 – 2014) of full-text content with 11.5 million words (McClure, 2014). 138,476 words were visualised as a 'conceptual atlas' that presented the forum discussion topics (see Figure 2.6).

¹⁵ The online forum *Humanist* is also a place where DH scholarly communications are carried out. In 1987, McCarty founded the long-standing email listserv *Humanist* as an international seminar on digital humanities (Rockwell and Sinclair, 2012; Nyhan, 2016). It offers a 'lasting, warm, indoor' place for DH scholars to discuss the intellectual, scholarly, pedagogical, and social issues that are related to DH. Also, it is a publication of the Alliance of Digital Humanities Organizations (ADHO) and an affiliated publication of the American Council of Learned Societies (ACLS). Although the list is mainly English, it provides rich data for analysing and understanding the history and scope of DH.



Figure 2.6: The screenshot of the topic network visualisation based on data from the Humanist discussion mailing list (1987 - 2014) (McClure, 2014).

McClure's chronological visualisation (Figure 2.6) showed that topical words were concentrated on hardware and software during the mid-80s (e.g., 'mainframe', 'microcomputer', 'workstation', 'wordperfect', 'printer', 'macintosh'), and the focus turned to the general growth in the 90s with many popular place names (e.g., 'Philadelphia', 'Pittsburgh', 'Pennsylvania', 'Georgetown', 'Quebec', 'Rutgers', 'Ottawa', 'Lancashire'). While the year 2000 marked the beginning of the DH disciplinary development and administration construction (e.g., 'dissemination', 'evaluation', 'preservation', 'speakers', 'invited', 'lecturer', 'workshop', 'organised'), the unique characteristics of DH that separated the field from other mainstream Humanities started to emerge from 2010 (e.g., 'collaborative', 'team', 'alliance', 'intersection', 'technologists', 'interdisciplinary'). Moreover, from the year 2011 and onwards, more diverse topics and different subjects and media appeared (e.g., 'GIS', 'Twitter', 'Gmail', 'blogpost').

The exhaustive forum analysis of McClure raised an important point that the subject of a field changes along with time. It is not only the dataset scale, data source, analysis method that affect the subject distribution, but also time, as an essential variant.

Accordingly, new questions emerged, such as, what type of data source can better represent the DH subject (e.g., journal, conference, Twitter, or blog)? Which data analysis method can better visualise the DH subject (e.g., social network analysis or topic modelling)? What time range can better cover the DH history and how to divide it into different periods? These questions require new empirical analysis, and we cannot simply answer them by reviewing existing studies, for each study has its own distinct dataset, method, and focus that are difficult to compare directly.

2.2 DH social actors

Studying DH scholars helps us learn the community they form. As a DH scholar, one could publish journal articles to form formal communication links with other co-authors while having continuous informal discussions and debates with other scholars on social media at the same time (Ross et al., 2011). By exploring how DH scholars interact with each other, one could also analyse the values that keep the community together, and the characteristics that differentiate DH from other disciplines. At the time of writing, few empirical studies put their focus solely on scholar identification. Most empirical studies that identified the DH community only saw it as a prerequisite of their process to explore the community activity and behaviour.

For example, Rockwell and Sinclair identified the names of DH scholars as part of the DH forum analysis, and 21 scholars were identified and visualised on the network (see Figure 2.7) (Rockwell and Sinclair, 2016).

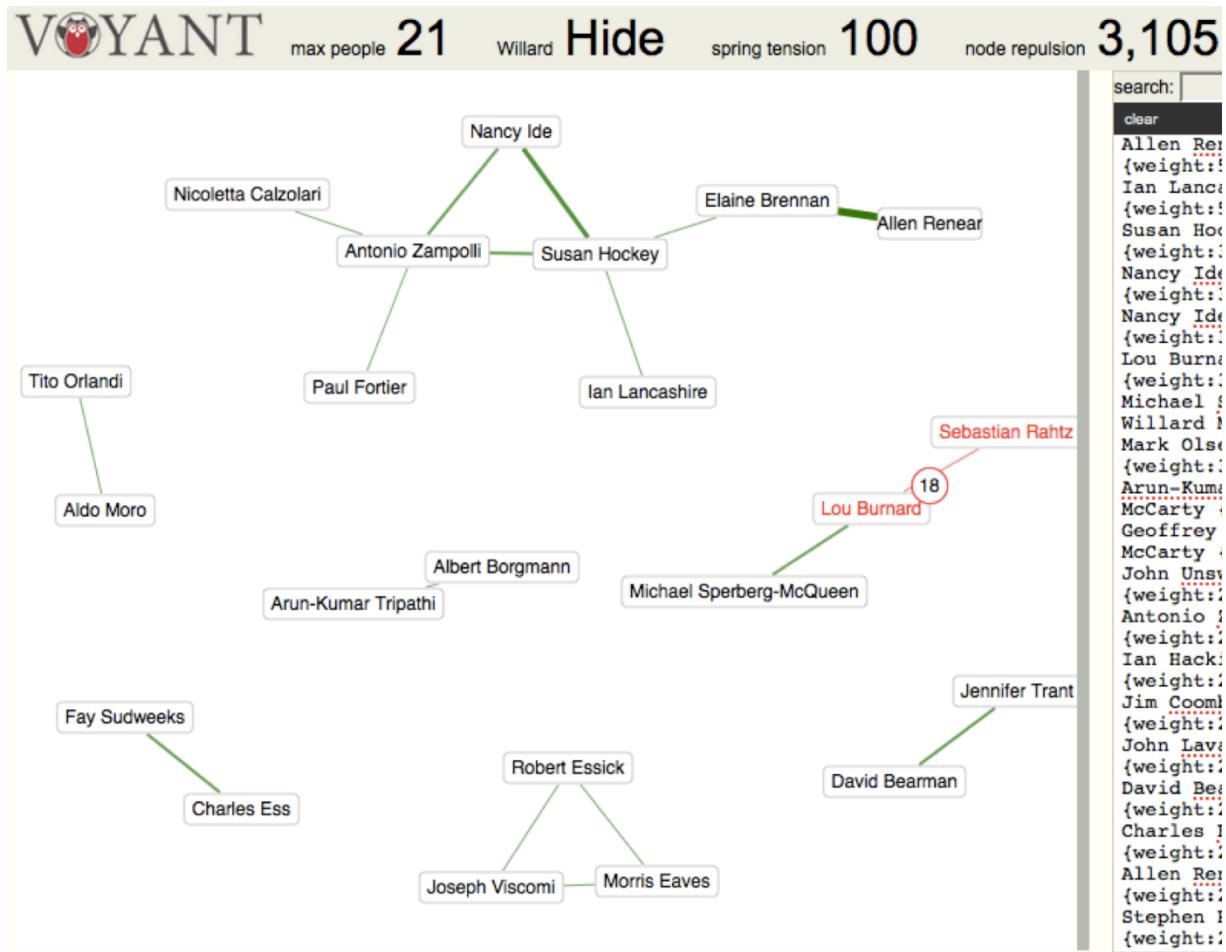


Figure 2.7: RezoViz view of the scholar co-occurrence network based on the Humanist Discussion Group listserv archives (Rockwell and Sinclair, 2016).

The edges in Figure 2.7 showed various reasons that people were connected (e.g., scholarly connection, regional link, co-authorship), but the sample size was very limited in representing the field, and Rockwell and Sinclair themselves described this study as only a 'swiftly fly-through' analysis.

This study, too, sees the process of identifying scholars as part of the research premise. This is because that there is no definite answer to 'who's in and who's out', as mentioned earlier (section 1.2.2 Scholar). The answer changes according to different recognitions, regions, fields, and datasets. Simply identifying names of scholars would not only be unhelpful for enriching our understanding of DH, but would also raise disagreements, debates and critiques about the community representation. Identifying and collecting personal data also brings challenges for data protection.

Therefore, to address the research question of ‘who has contributed to the development of DH’, simply identifying the membership of the DH community is not practical nor pragmatic. Instead, investigating the collaboration and communication patterns of DH scholars could help us learn more. This section reviews previous efforts made to quantify these two different types of DH scholarly communication (i.e., 2.2.1 – Co-authorship, 2.2.2 – Twitter activity).

2.2.1 Co-authorship

As reviewed in 2.1 DH subject specialty, the topic of collaboration is an important subject that is closely related to DH. It was highlighted in many DH studies, and seemingly a characteristic that helped to shape the DH community (Bradley, 2016; Moravec, 2018). For example, Deegan and McCarty stated its importance:

Collaboration within digital humanities is both a pertinent and a pressing topic as the traditional mode of the humanist, working alone in his or her study, is supplemented by explicitly co-operative, interdependent and collaborative research. This is particularly true where computational methods are employed in large-scale digital humanities projects (Deegan and McCarty, 2012, p. i).

Many studies that have researched scholarly collaboration have used co-authorship method to assist their analysis (e.g., Cronin, 2005; Fagan et al., 2018; Köseoglu et al., 2018), and DH is no exception (e.g., Nyhan and Duke-Williams, 2014a; Tang et al., 2017). This section will review existing DH co-authorship studies and bring a new focus of DH co-authorship and collaboration that helps to form this thesis.

In order to find whether the collaborative nature was more frequent in DH or in the mainstream humanities, Spiro applied statistical methods to analyse DH journal publications (Spiro, 2009). She collected 259 papers from a humanities journal (*American Literary History, ALH*) and 145 papers from a DH journal (*Literary and Linguistic Computing, LLC*) 2004 – 2008, and found that only 1.93% of the articles in the former had more than one author while 48.28% of the articles in the latter had multiple authors.

Spiro interpreted the cause of this significant difference as relating to the distinct requirements of academic practices between DH and the Humanities. DH projects, for example, require more collaborations, multiple techniques from different kinds of

specialists, various sources and equipment, while the mainstream humanities studies (e.g., literature studies) can be conducted by a lone author who is doing reading, writing, and thinking independently.

In 2014, Nyhan and Duke-Williams uncovered a surprising result by comparing co-authorship patterns between DH and Geography publications. More specifically, they collected 2,291 articles in *CHum* (1966–2004) and *LLC* (1986–2011), and compared the data with a Geography journal, the *Annals of the Association of American Geographers* (AAAG) (1966–2013). By analysing the co-authorship pattern, they found that the increase in co-authored papers in AAAG is statistically more significant compared to that in *CHum* and *LLC*. They noted that not only was the increase of co-authored papers more significant in AAAG but that such an increase had been found in most disciplines during recent decades. According to their results, DH as an emerging field, based on co-authorship analysis, is not as unusual as most people assumed.

As we know from the previous section, different data sources might present different DH subject structure; would a change of data source also affect the co-authorship pattern in DH? Weingart's continuous studies on the DH conference demonstrated that this might not be the case in terms of co-authorship. Instead of using journals, Weingart analysed the co-authorship structure at ADHO conferences. In his studies, the co-authored papers stayed at around 60% of the total from DH2013 to DH2016 (Weingart, 2013a, 2013a, 2014a, 2015a, 2016b), although the number of submissions and participants varied from year to year. Even at DH2015 Australia, which had affected the distribution of conference topics more significantly than any other ADHO conference, the co-authorship proportion remained at around 60% (Weingart, 2014a). This figure is similar to that of journal publications within the similar time frames in *LLC*. For example, in Nyhan and Duke-Williams' results, the multi-authored papers accounted for 54% of the total in 2009 and 53% in 2010 (Nyhan and Duke-Williams, 2014a, p. 392). This range of percentage matches the number of some regional conferences, too. For example, the Dutch-speaking DH conference DHBenelux had around 58% papers that were multi-authored in 2016; this number increased to around 60% in 2017 and 66% in 2018 (Kemman, 2016b, 2017, 2018). Although DHBenelux

distinguishes itself from the large-scale ADHO conference, its co-authorship pattern was the same as that of the ADHO conference and recent *LLC* publications.

Nevertheless, other regional conferences showed a different pattern. For example, the DH conference in Taiwan (DADH 2009-2012) had around 40% co-authored papers (Chen and Hsueh, 2013), while the German DH conference (DHd2018) had around 72% (Henny-Krahmer and Sahle, 2018). The difference in the co-authorship proportion might be due to a focus on different conference topics. For example, studies have shown that the DADH conference focused mostly on humanities topics (mainly historical studies) (Chen and Hsueh, 2013), while the DHd conference appeared to have more diverse topics (Henny-Krahmer and Sahle, 2018). The difference could also be due to their languages. As Flanders once indicated 'the collaboration of conversation is predicted on the norms of language' (Flanders, 2016); would scholars co-author differently in distinct language contexts? For example, an empirical study conducted by Larivière et al., found that language and geographical proximity influenced the choice of collaborators in the social sciences and humanities (Larivière et al., 2006).

Although many people do not co-publish, the studies reviewed above suggest that DH might have a small but strong set of active co-publishers. Nyhan and Duke-Williams found that 'a relatively small cadre of authors who co-publish with a wide set of other authors, and a longer tail of authors for whom co-publishing is less common' (Nyhan and Duke-Williams, 2014a, p. 396). Later, a group of scholars led by De la Cruz constructed a co-authorship network based on 178 *DHQ* articles 2007 – 2014, and also found similar results (De la Cruz et al., 2015). The largest cluster of their co-authorship network only covers 16% of the authors that have published in *DHQ*. Furthermore, in Tang's study, Tang and his collaborators also agreed with the idea of a strong but small group of active scholars in DH, and they indicated that DH had a 'highly fragmented' co-authorship network:

The co-authorship network was shown to be highly fragmented, consisting of numerous small components that resemble the 'plural worlds' model, without an extensive giant component often observed in neighbouring fields, such as digital libraries, for example (Liu et al. 2008). The clustering coefficient is very

high, indicating that collaborators tends to form closely-knitted clusters. (Tang et al., 2017)

In other words, based on the above studies, the DH co-authorship remained ‘highly fragmented’, while at the same time, ‘closely-knitted’ as a ‘small world’. It seems contradictory, but it is understandable. The highly connected small world might be but one component of ‘plural worlds’ where different fields are involved under the ‘big tent’ of DH (Lagemann, 1989; Tang et al., 2017, p. 987). Individual, highly connected small worlds could be brought together while remaining isolated from each other. Under such an assumption, this argument might imply that DH subjects have not been well-integrated, and the DH groups are not fully connected (whether loosely or closely) (Tang et al., 2017). Still, we see the DH community and events keep consolidating across the world (Gil and Ortega, 2016), and this may be why people started to describe DH as a ‘trading zone’ or ‘meeting place’ while the ‘big tent’ and ‘archipelago’ notions are no longer appropriate (McCarty, 2006; Svensson, 2012). If still fragmented, how do these small worlds distribute? Would geographic and language factors influence such distribution?

To address these questions, further empirical analysis is required. It is one of the research aims of this study to explore the DH co-authorship publishing pattern and the factors that influence it. Further discussion on methodology will be presented in the next chapter (chapter 3 Methodology).

2.2.2 Twitter activity

Formal publications offer one view of the DH community (Leydesdorff and Salah, 2010, p. 34). Empirical studies on social media can extend this view to include more of the social aspects of DH (Witting, 2018). Sociometric methods have been applied to a variety of fields, such as interpersonal relationship studies (Jones et al., 2012), group collaboration (Kim et al., 2008), literature studies (Anheier et al., 1995), as well as the community studies of DH (Quan-Haase et al., 2015a; Grandjean, 2016).

Twitter, among other social media, is one of the most popular microblogging platforms that not only supports regular social activities at an individual level, but also enhances and expands the ‘backchannel’ communications among academics, despite issues such as rising distractions (Ross et al., 2011). Even though not all scholars use Twitter

(Van Noorden, 2014), it has been accepted by more and more academics since its launch in 2006 (Collins et al., 2016). Many Twitter users in academia have found it useful in terms of promoting their work and ideas, virtually participating in conferences, and following research-related news and discussions, such as new academic publications, funding opportunities, conference announcements, and science policies (Côté and Darling, 2018).

Despite DH's early adoption and active use of Twitter (Ross et al., 2011, p. 229), only a few empirical studies have made the attempt to examine and classify different types of Twitter usage by the DH community. Their results revealed two main purposes of DH scholar using Twitter – 'information dissemination' and 'social networking'. These two activities were also called 'writing' and 'chatting' by French (French, 2009).

In 2009, French found that twice as many people tweeted about THATcamp than actually attended it (French, 2009). She also emphasised that Twitter was more about 'writing' than 'chatting' where news about the academic events were shared to the audience.

This result is consistent with the results of Twitter user studies across other disciplines. In 2010, Kwak and his collaborators became 'the first to look at the entire Twittersphere' (Kwak et al., 2010, p. 599). Based on 41.7 million user profiles, they investigated user participation in trending topics via online interactions. One of their main results showed that most Twitter connections were one-way rather than bilateral interactions, and therefore the average Twitter usage was more about 'information dissemination' rather than 'social networking'.

Quan-Haase et al.'s study also agreed with it (Quan-Haase et al., 2015a). They conducted semi-structured interviews with 25 DH scholars and studied their ways of using Twitter. Their results showed that their interviewees used Twitter more often for informational purposes rather than for social networking (more followed organisational accounts than personal accounts).

Holmberg and Thelwall also studied such one-way information sharing activity and found that Twitter was used statistically more often by experienced scholars with a higher position in the academic hierarchy than by newcomers (Holmberg and Thelwall,

2014). Thus, it is not difficult to understand why the majority of users chose to share or subscribe information instead of having conversations across different hierarchies.

However, Holmberg and Thelwall also agreed with others who had different opinions arguing that Twitter's role of being an information network would later change to a social network as users became more selective and experienced (Myers et al., 2014, p. 498). They compared the Twitter usage of DH scholars to that of scholars from 9 other disciplines and found that users in DH had more conversations and discussions (38% of the tweets) than those in other disciplines (astrophysics, biochemistry, economics, history of science, cheminformatics, cognitive science, drug discovery, social network analysis, and sociology), whereas scholars in sociology and the history of science seemed to have little conversational activity on Twitter (Holmberg and Thelwall, 2014). In DH, only 15.5% of the tweets included links (an indicator of information sharing), and this figure was relatively low compared to the other nine disciplines (e.g., 75% in astrophysics). These results suggested that for DH scholars, Twitter was used more as a platform of social networking.

Yet, these two purposes for using Twitter are not as contradictory as they were described. User activity is a complicated system consisting of numerous kinds of interactions, and binary classification is not useful when addressing such complex factors.

Ross and her co-authors suggested that there might be more purposes when using Twitter than just 'writing' and 'chatting'. They suggested the concept of Twitter being used as a 'backchannel' for the DH community, particularly during conferences and academic events, that embraced information dissemination and social networking, as well as building communities and values. By analysing 326 users during three DH conferences from June to September 2009, they classified 4,574 tweets into seven types (comments on presentations; sharing resources; discussions and conversations; jotting down notes; establishing an online presence; asking organizational questions; and unknown).

Grandjean conducted a series of network studies to visualise the DH community on Twitter (Grandjean, 2013a, 2013b, 2014, 2015, 2016), from which he found that the DH community was rather a 'small world'. This result agrees with the co-authorship

studies reviewed above that there is a cohort of DH scholars that are closely connected (e.g., Nyhan and Duke-Williams, 2014a; Tang et al., 2017).

In one of Grandjean's works, he studied the 'social networking' activities of DH users by analysing their patterns of mentioning other users. At DH2014 in Lausanne, Switzerland, Grandjean and Rochat collected 16,903 tweets from almost 2,000 users who posted with the hashtag (#DH2014), and constructed a 'mentioning network' of DH scholars (Grandjean, 2014; Rochat, 2014), (see Figure 2.8).

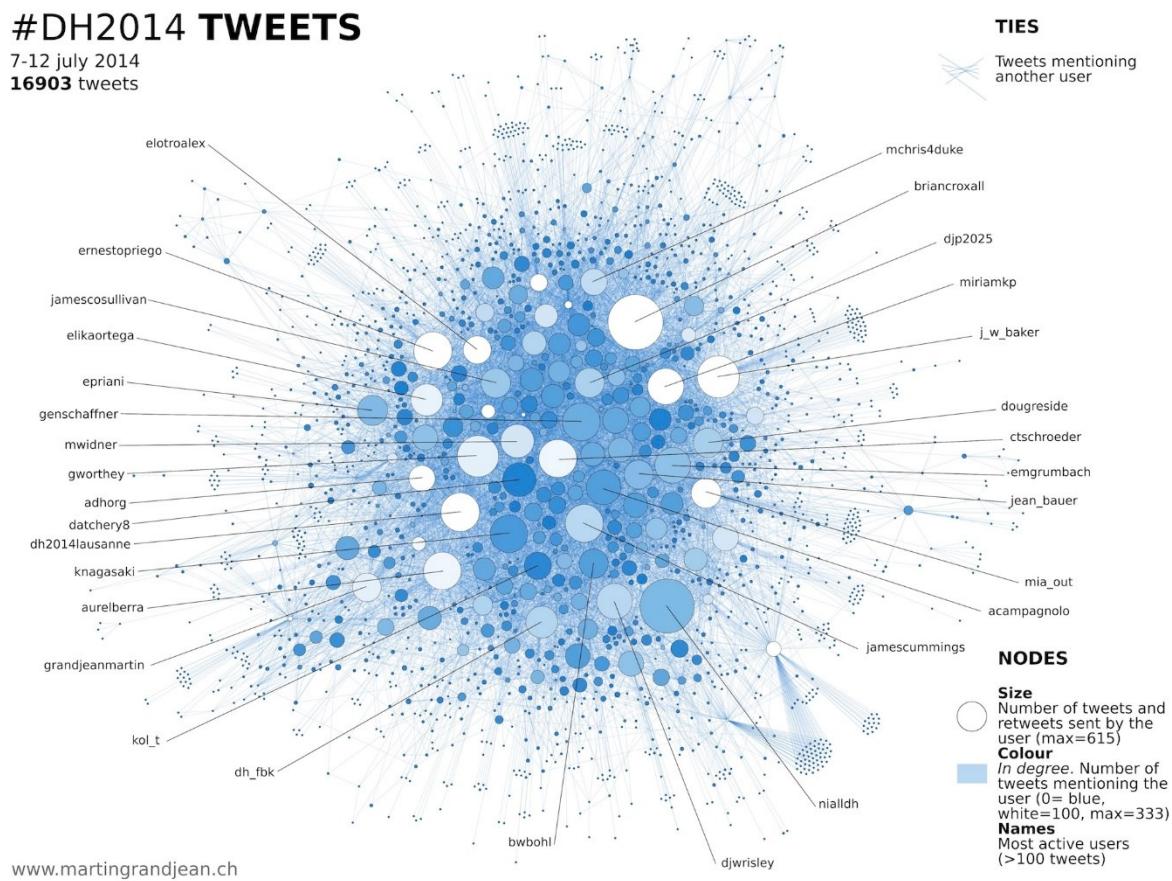


Figure 2.8: Network visualisation of interactions on Twitter during #DH2014 by (Grandjean, 2014; Rochat, 2014). Use in this thesis has been permitted by the author.

As shown in Figure 2.8, the nodes were weighted according to the number of tweets the user posted, and their colours depended on the number of mentions they received (the more mentions the whiter). Although there was no direct co-relation between the most active users (who posted more tweets) and the most mentioned users, they found that the network was very dense, and the keynote speakers were mentioned heavily. This is a typical structure of a 'small world' where most nodes can be reached

from every other node by a small number of steps. Considering that DH2014 was the largest DH conference in the year with 750 attendees from all around the world, this ‘small world’ representation was unexpected. Did these 750 attendees from different countries across the world all know each other very well? The answer would seem to be ‘no’. How then did they form a closely connected ‘small world’ network? Even though the co-authorship studies reviewed above demonstrate a similar result (i.e., small world), they were based on a distinct dataset with distinct relationships (i.e., co-authorship on publications) (e.g., Nyhan and Duke-Williams, 2014a; Tang et al., 2017). This might be partly due to the fact that the 750 participants were not all on Twitter engaging with the hashtag #DH2014. The people who attended the conference might not have been the people that tweeted on Twitter. As we know, the ‘mentioning network’ (Figure 2.8) contains around 2,000 Twitter users, while there were 750 participants physically presented at the conference. Although there is, supposedly, an overlap between these two groups, such an overlap might be just a small cohort of people that know each other well and form a close-knitted network. Other empirical studies also found that there is little overlap between scholars producing publications and scholars tweeting on social media. For example, Holmberg and Thelwall tried to identify the most productive authors in 10 disciplines (including DH) on Twitter, but only 1 out of 20 scholars can be found (Holmberg and Thelwall, 2014).

In addition, this mentioning network constructed by Grandjean was based on data that was collected specifically from a conference (with hashtag #DH2014) during the time that it was held. It would be expected to find tweets mentioning the same people (e.g., keynote speaker) and talking about the same things (e.g., presented papers). Because of these similarly mentioned users and topics, it was no surprise to find that all the users on the network were connected closely and that they formed a dense network of mentioning relationships.

A follow-up study constructed another dataset containing DH Twitter users and tweets by searching keywords among user bios (i.e., profiles) (Grandjean, 2016). The visualised network was based on a new relationship – ‘following’ – instead of ‘mentioning’. Compared to ‘mentioning’ activity, ‘following’ brings more insights about how people subscribe to their information source and how they use Twitter for ‘information dissemination’ purposes. Mentioning activity, on the other hand, reveals

more about how people carry out conversations for the purpose of ‘social networking’. More specifically, he selected 2,538 Twitter accounts and retrieved the ‘following’ connections among these accounts (Grandjean, 2016, pp. 3–4), and from these he visualised the network using Gephi (see Figure 2.9).

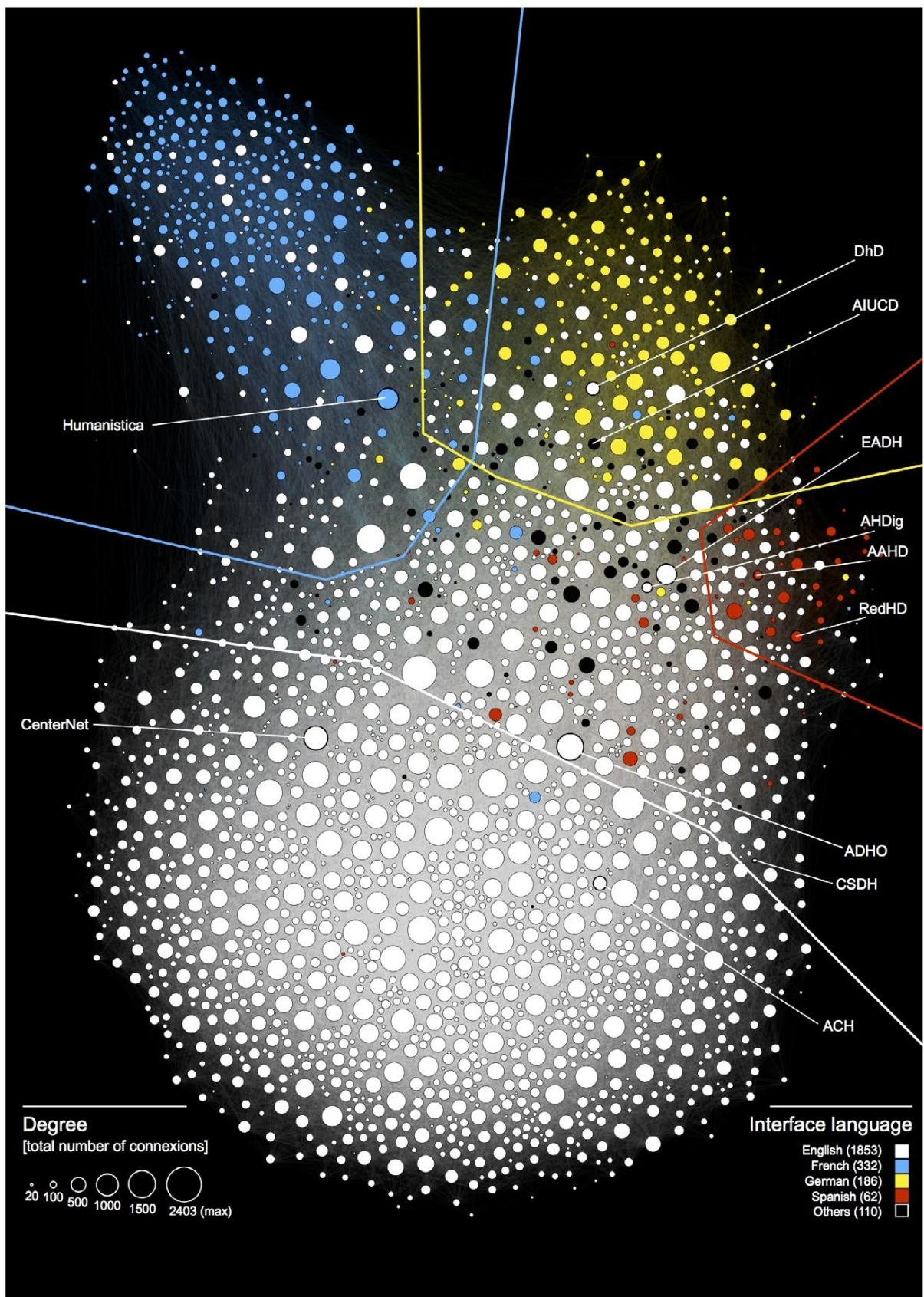


Figure 2.9: Digital humanities network on Twitter: 2,538 users following each other colour-coded with language. Use in this thesis is under Open Access CC 4.0 licence.

Although his results still showed a dense DH social network, this time there were subgroups closely connected, such as French-speaking and German-speaking clusters (as Grandjean colour-coded on Figure 2.9). The new findings showed that language seemed to be an essential element that affected the grouping of the DH social network (Grandjean, 2016, p. 1). Grandjean also pointed to the geographical and cultural factors that were closely related to the node distribution on the network, particularly the French-speaking cluster that has emerged in recent years. It is similar to what has been discussed earlier (section 2.2.1 Co-authorship) that language and geographic factors influence how people communicate as well as collaborate, and these key factors will be discussed in the next section (2.3 DH information use environment).

Another similarity with the co-authorship study is that both the Twitter network and the co-authorship network pictured the DH community as a ‘small world’. Do these two networks represent the same ‘world’? As we have seen, the two networks are made of potentially different groups of people (i.e., authors on publications and users on Twitter). Moreover, we could detect different patterns of cluster distributions from them. The network of co-authored scholars is only one isolated small cluster; there are a large number of authors and co-authors of DH publications who are not in that co-authorship network but form their own even smaller co-authorship links and are mostly disconnected. Thus, the whole co-authorship picture should be a ‘plural worlds’ model consisting of numerous separated components. On the Twitter network, most identified users are shown in the network, so there are hardly any isolated small groups. Clusters detected on the Twitter network are closely connected to each other and, thus, form a dense ‘small world’ that is clearly different to the co-authorship network.

Nevertheless, we need to be aware that we are comparing two different kinds of network studies of different scales and by different people at different times. They can be compared directly only to a limited degree. New approaches, accordingly, are needed to find the community both in publications and on social media to support more comparable research.

In addition, because Twitter, by its nature, holds short, fragmentary and abbreviated messages that are not suitable for normal text analysis methods, the need for innovative Twitter-based approaches is growing. Further comparative study with an

empirical approach is needed, and this is one of the objectives of the current study that will be discussed in chapter 3 Methodology.

2.3 DH information use environment (IUE)

Recent years have witnessed increasing attention being paid to diversity trends and movements (e.g., #transformDH). Nevertheless, empirical studies have shown that the values of DH and the ways in which they are practiced in real life appear to be different. This section reviews previous efforts made to quantify two IUE factors (i.e., 2.3.1 – Gender, 2.3.2 – Affiliated country), and discusses their achievements as well as challenges and problems, and how we could further solve the research question.

2.3.1 Gender

Gender is a socially-constructed classification system that contains multiple properties differentiating between masculinity and femininity. Depending on the context, gender can be used to describe different models, such as recognised social role, or gender identity (Haig, 2004, pp. 1945–2001). The division of people into two mutually exclusive genders is known as ‘gender binary’, and it has been a classification standard not only in Western cultures, but also in most cultures across the world (Bray, 2007; Nadal, 2017). There are many other models that are usually labelled as ‘non-binary’ or ‘genderqueer’, and apart from ‘masculine’ and ‘feminine’, they have other descriptors, e.g., ‘transsexual’, ‘hijra’, ‘intersex’ (Hill and Willoughby, 2005). Although there is much to be celebrated about increasing gender variance from the primary ‘masculine’ and ‘feminine’ model, to the best of the candidate’s knowledge, most existing empirical gender studies in DH have applied gender binary approaches. It is not a surprising finding as identifying and collecting quantitative data of transgender scholars in DH is not feasible within the current means, e.g., the level of resolution not captured in most datasets. Therefore, this current study will mainly focus on researching and discussing gender questions in DH using the binary model (with ‘other/unknown’ where appropriate), despite the candidate’s full awareness and recognition that there are scholars in DH who are gender diverse.

There are few gender studies in DH that have applied empirical approaches. Rather than author groups on publication, DH users on Twitter were the first group of research subjects to be analysed. In 2010, by retrieving gender data of 164 DH Twitter users

via Twitteranalyzer, Fluharty found that only 40% (65 out of 164) of them were female (Fluharty, 2010). He pointed out that this figure was unexpected given that DH was closely related to the Humanities where women normally accounted for the largest part (e.g., 80% librarians, 66% of archivists, 66% public historians), but he also explained that this figure matched the general gender ratio of all users on Twitter in 2009 (with 43% female users) (Nielsen Mobile, 2009).

Additionally, Fluharty found that female Twitter users in DH tended to have fewer followers (average 564) than the average number (779), while the male users tended to be 'a little too selective' in the users they followed, with a following-to-follower ratio of 0.57 while their female counterparts had 0.69. This male elite attitude is not only found in DH, but also found in other fields. For example, Artwick found that the tweets of men were much more likely to be retweeted and disseminated than those of women (Artwick, 2014). Moreover, in her dataset of 2,733 tweets, male reporters from large newspapers had not quoted a single tweet posted by female reporters, although this did not mean that there were no women's voices in their Twitter streams. From these results, it seems that tweets by male users in these groups were more likely to be disseminated and subscribed to, and so it is not difficult to infer that men would have a lower following-to-follower ratio (i.e., fewer following and more followers) than women. On the other hand, some studies also pointed out that women on Twitter are significantly more open and transparent than men, while there is little difference in both groups' Twitter presence, topics, opining or gatekeeping (Lasorsa, 2012). In Blumell's study, female users tended to be 'softer' than their male counterparts, revealed more about their jobs, personal lives and everyday activities, and they linked more to external websites (Blumell, 2019). This might also suggest that women are more likely to follow back their followers and are less likely to hold an elitist attitude. However, such gender studies are not found in DH. A new study is needed to investigate the gender difference in the DH Twitter community and how these groups use Twitter in different ways for communication and collaboration.

Compared to research on Twitter users, DH gender studies on publications started rather late. In 2014, Weingart analysed the conference dataset containing author names, affiliations, and keywords from previous ADHO conference programmes, and manually assigned gender information for each author as 'male, female,

unknown/other' (Weingart, 2014c). He pointed to the unbalanced gender ratio of attendees at DH conferences (around 30% female scholars each year from DH2004 to DH2013).

However, this percentage does not seem to be the case at regional DH conferences. For example, Tello analysed the Spanish-speaking DH conference (HDH2015, *Humanidades Digitales Hispánicas*) (Tello, 2015a). Among all 229 attendees, there were around 55% females, and as Tello discussed in the blog, it indicated that women were leading the field there which was very different to the ADHO conferences:

I think the DH field in Spanish language is doing a good job against gender discrimination. One female president abandons the charge and another female president took up. Another great example of that was the panel were 4 women presented their work of creating groups and networks around DH in Spain and Latin America. [...] What I mentioned in my last post about women in leading positions in DH field, is also truth seeing the data of speakers. (Tello, 2015b)

However, as discussed earlier at section 2.1.2 (Conference), the distribution of topics and authors at conferences is more unstable and more likely to change compared to those of journal publications. Thus, it is difficult to examine the patterns of demography in conference attendance.

Apart from conference attendance, Weingart combined his gender dataset with topical keywords (2013 – 2015), and found that certain subject matters were gendered at DH conferences (Weingart, 2016a). Apart from a slight bias in subjects that women are more likely to present, Weingart's study seemed to point to a less obvious bias against female scholars:

Women are twice as likely to use the 'Gender Studies' tag as male authors, whereas men are twice as likely to use the 'Asian Studies' tag as female authors. Subjects related to pedagogy, creative / performing arts, art history, cultural studies, GLAM (galleries, libraries, archives, museums), DH institutional support, and project design/organization/management are more likely to be presented by women. Men, on the other hand, are more likely to write about standards & interoperability, the history of DH, programming, scholarly editing, stylistics, linguistics, network analysis, and natural language processing / text analysis. It seems DH topics have inherited the usual gender

skews associated with the disciplines in which those topics originate. (Weingart, 2016a)

To the best of the candidate's knowledge, at the time of writing, no studies have analysed the gender of DH authors of journals and books. It is an important question, and one which will help our understanding of 'who we are'. The current study will, thus, investigate the gender difference within the DH community on journal publications, as well as on Twitter. Further research design will be discussed in chapter 3 Methodology.

Although there are no existing studies of gender and DH publication patterns, reviewing previous studies in other fields should help to inform the current research. For example, many studies found that female scholars tend to have fewer publications than men. Based on a review of several studies, Larivière and other authors showed that women tend to publish around 70% to 80% as many publications as men (Larivière et al., 2011). Later, Rørstad and Aksnes arrived at similar percentages in their large-scale survey on 12,000 scholars, although they also noted significant variations across fields and academic positions (Rørstad and Aksnes, 2015, p. 325). They found that for all levels of academic positions, male scholars had a slightly higher publication rate than female scholars (on average 0.25 publication per person per year), and the gap increased by age (Rørstad and Aksnes, 2015, p. 326).

As mentioned above (see 1.2.3 Environment), recent years have witnessed increasing attention being paid to DH feminism, and there are growing numbers of related journal issues, books, and conference panels that have boosted the conversations about its significance. Would the gender publishing pattern of DH scholars show something exceptional compared to other disciplines? New questions related to gender and DH publications need to be addressed.

2.3.2 Affiliated country

Despite advances in the provision of, and access to, digital communication technologies, affiliated location remains a key determinant that influences the scholarly practice. Pan et al. conducted a large-scale study that investigated the relationship between academic collaboration and scholars' affiliated location (Pan et al., 2012). They not only found a strong correlation between authors' affiliation and their collaboration strengths (i.e., the closer the affiliated locations, the more collaboration

between scholars), but also discovered a linear growth between national funding and a country's research impact (i.e., the more research funding a country grants, the more significant the research impact). This means that affiliated country (or affiliated location) still plays an important part in academic communication and collaboration.

DH is not an exception. As mentioned earlier (e.g., 2.2.1 Co-authorship and 2.2.2 Twitter activity), we found that in DH, geographic location and language not only influence how people communicate informally on social media but also how they collaborate and produce formal research in publications (e.g., co-author). Therefore, understanding the geographic distribution of DH scholars helps us see the field's global landscape and structure. Some empirical studies have tried to answer questions that are related to the geographic distribution of DH scholars.

In 2006, Terras made the first attempt to count the countries (Terras, 2006). Based on the programmes of 10-years of ADHO (ACH/ALLC at that time) conferences (1996-2005), she found that there was a large group of presenters from the USA and Canada (around 37% and 24% respectively), and all the presenters were from Western countries with the distinct absence of China and India. Later, in 2012, Terras provided another snapshot review of the field (see Figure 2.10) using data collected from centerNet that confirmed her finding (Terras, 2012a).

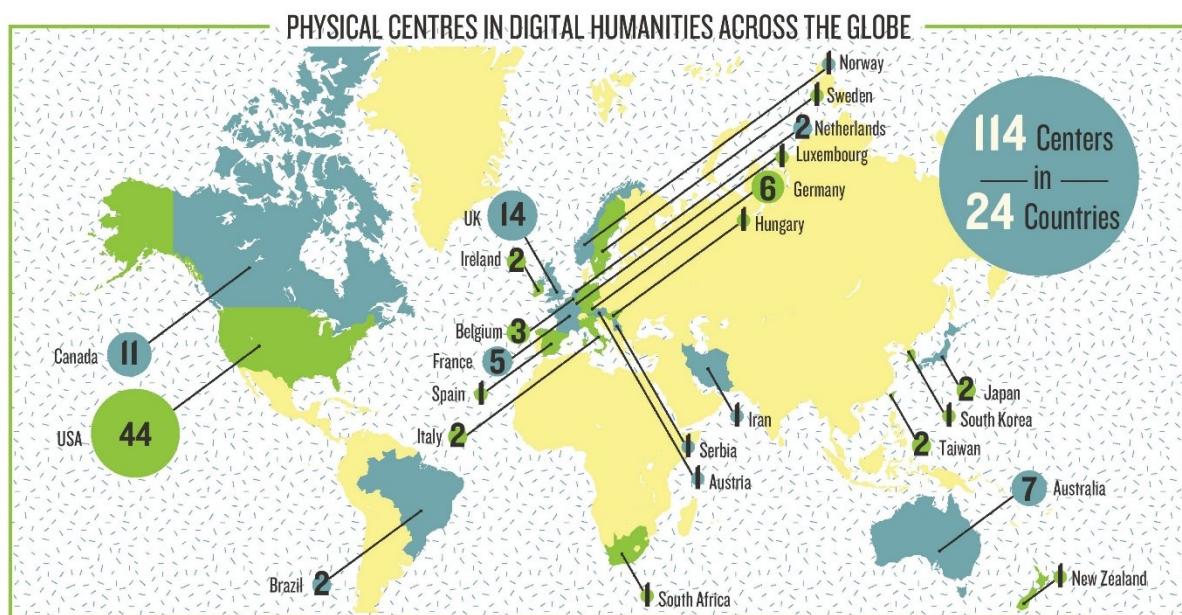


Figure 2.10: Physical centres in DH across the globe by (Terras, 2012a).

Although her infographic witnessed the increasing growth of DH in many directions, the global map has many blank spaces (Risam, 2018) and possible omissions (Mahony, 2018). Such distribution can be found in many DH geographic studies. For example, in the same year, Clavert analysed the country distribution of DH2012 reviewers and demonstrated that the majority of them were from the UK and the USA (Clavert, 2012).

As mentioned, DH2015 in Australia attracted the most diverse geographic distribution of scholars compared to its previous years; the number of scholars from Asia almost doubled in proportion while delegates from Oceania were seven times greater, as shown in Figure 2.11 (Weingart, 2014a).

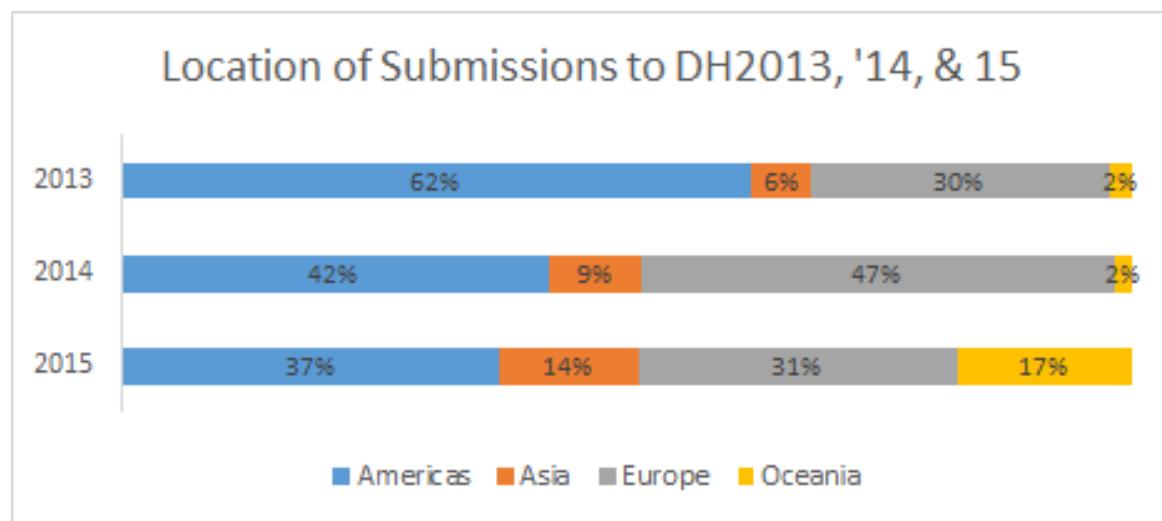


Figure 2.11: Location of submissions to DH2013, DH2014, and DH2015 (Weingart, 2014c). Use in this thesis has been permitted by the author.

Studies on DH journal articles have also shown that the author pool consisted of predominantly Western scholars. A group of scholars led by De la Cruz visualised a colour-coded co-authorship network based on 178 *DHQ* articles 2007 – 2014 (De la Cruz et al., 2015). They identified 170 unique authors, and around 73% of them were from North American institutions and 26% were affiliated in Europe (see Figure 2.12). This means almost no authors outside of North America and Europe published with *DHQ* from 2007 to 2014.

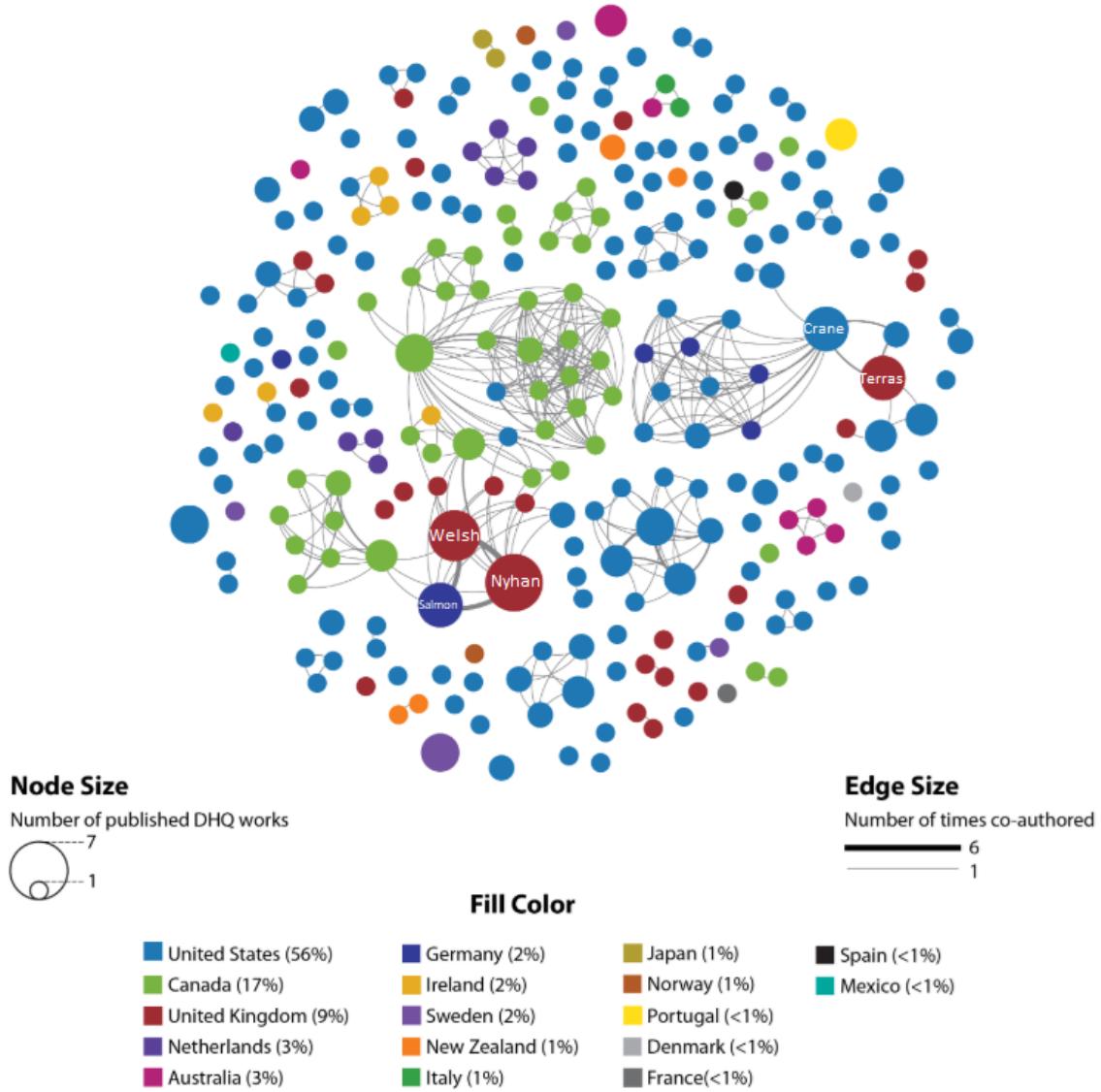


Figure 2.12: Co-author network for the journal Digital Humanities Quarterly (2007 – 2014) by (De la Cruz et al., 2015). Use in this thesis is under Open Access CC 4.0 licence

Additionally, geography as a key influence is particularly effective and powerful in regional conferences. For example, most participants at the Taiwan DH conferences (DADH2009 – 2012) were Chinese and Japanese (Chen and Hsueh, 2013), while 80% of presenters at the German conference (DHd2016) were affiliated with a German institution (Tello, 2016). At the DHBenelux conferences (2016 – 2018), more than 90% of authors were from the Benelux region (the Netherlands, Belgium, and Luxembourg) (Kemman, 2016a), while at the DHN conferences (2016 – 2018), almost all scholars were based in Nordic countries (Mäkelä and Tolonen, 2018). The result is expected, as these conferences are language and region-focused. DH has been practiced in

different ways and in different linguistic-cultural contexts, although the situation is gradually changing with the international growth of DH (Fiormonte, 2014, p. 3).

From the studies reviewed above, it is not difficult to notice the geo-linguistic influence in the formation of different DH communities across the world. Nevertheless, for relevant empirical research, the most urgent problem to be solved is how to find a comprehensive data source and to construct an inclusive dataset to represent the entire DH community. Such problems are raised by Fiormonte, for example:

[...] the different nuances of the linguistic-cultural problem, cross-cultural representation within the international organizations of DH, the consequences of the English-speaking dominance in the processes of discussion and factual evaluation, the relationship of DH to colonial and subaltern studies [...] (Fiormonte, 2014, p. 2)

In addition, DH scholars who are from countries that are not connected to or recognised by ADHO are underrepresented. O'Donnell et al. offered one explanation about the gap between high-income and low-income economies:

[...] our international and collaborative activity is conducted along a primarily east-west axis among a relatively small number of mostly contiguous high-income economies in the northern hemisphere: Japan, Taiwan, South Korea, Canada, the United States, the countries of western and central Europe, and, in the South, Australia and New Zealand. (O'Donnell et al., 2015)

Many scholars have drawn attention to the DH development in the 'blank areas' on Terras' infographic map. For example, Galina pointed to the increasing significance of the Spanish-speaking community and their Twitter profile building (e.g., @Red_HD) (Galina, 2014, pp. 312–313). She emphasised the importance of diversity, regions and languages other than English within the DH context. Mahony also pointed to the rapid DH growth in China, and that 'the anglophone world could do more to engage with practitioners and potential colleagues in this new vibrant and emerging area' (Mahony, 2018).

Few empirical studies have made an attempt to examine country distribution on a more inclusive and large-scale dataset. By expanding the dataset and comparing both publications and Twitter communities, this study revisits the geo-linguistic status of the

DH community, considers differences to previous research results, and how and where the differences lie.

2.4 Discussion

DH is undergoing substantial growth that can be studied from many data sources that are related to its research practice, social communication, and design of infrastructure.

Empirical approaches (e.g., bibliometric and sociometric analysis) have limitations and bring only partial representations, and thus, they can introduce potential biases to the results and alter our understandings of the field. Nevertheless, if applied appropriately and with the help of qualitative analysis and interpretation, quantitative analysis has the ability to build 'a radical shift in how we think about research' (Boyd and Crawford, 2012, p. 665), and offer practical generalisations about patterns within the data from macroscope and microscope views.

This chapter reviewed literature devoted to solving the three research questions shaped in chapter 1 under the structure of invisible college. The discussions demonstrated various patterns of DH subject and community based on different sources (e.g., DH events, journals, grants, social media, curricula) and different methods (e.g., statistical approaches, data visualisation, network analysis, topic modelling). In addition, new questions have emerged. How to collect sufficient data to represent a comprehensive DH network? How is DH both a 'small world' and 'plural worlds' at the same time? How might we identify the influences that formed the network?

By providing a systematic literature review that leads to the empirical experiments in later chapters, this PhD study learns from these previous efforts and builds on them by extending their work beyond any specific method but towards a refined research model that combines both quantitative and qualitative methods to systematically study the history, formation and ever-changing structure of DH. Instead of debates between 'small world' or 'plural worlds', this study provides solutions that link demographic factors from a broader context. Therefore, it enables us to review the history and disciplinary structure with the help of wider multidimensional elements such as historical, political, ethnographic and cultural influences.

The next chapter (3 Methodology) will introduce the research model and how it will be applied to answer the research questions systematically.

3 Methodology

This chapter introduces the refined ‘Invisible College’ model as the methodology of this thesis. Section 3.1 (Data Sources) introduces the reasons for choosing journal articles and Twitter as the data sources. Section 3.2 (Subject Specialty) discusses the methods to address the *subject* research question, while section 3.3 (Social Actors) presents the methods to investigate the *scholar* research question.

In total, this study builds four different networks, and each section has two networks. Section 3.2 (Subject Specialty) introduces *author co-citation network (formal)* and *Twitter hashtag co-occurrence network (informal)* to study the DH intellectual structures. Section 3.3 (Social Actors) proposes *author co-authorship network (formal)* and *Twitter user co-retweet network (informal)* to explore the DH scholarly networks. Section 3.4, Information Use Environment (IUE), introduces methods to further analyse the two social networks built in section 3.3. By adding gender, affiliated country, and language information to the two networks, IUE discusses the methods used to investigate the formation of the two networks. This chapter introduces the rationales for each of these four networks and how they can best answer the research questions proposed in chapter 1.

3.1 Data Sources

This study collects data from both formal and informal communication channels. As discussed in section 1.3 (Methodological framework), it constructs a bibliometric dataset based on data from journals (representing the formal DH communication channel) and a Twitter dataset (representing the informal DH communication channel). Scholarly communications have been carried out at a variety of levels and via different channels, such as the individual level, group level, and societal level, and sometimes, as this study acknowledges, it is difficult to differentiate the boundaries between ‘formal’ and ‘informal’. This study decides to analyse the two datasets mentioned above not only because they are suitable sources for scientific study and because of their compatibility with the well-grounded invisible college model, but also because they complement each other with different ways of communication and can thus advance our understanding of the DH communities beyond the homogenous perspective.

Bibliometric analysis is often seen as a ‘rear-view mirror’ that can help to reveal the knowledge map of a field through its formal publications. Analysing bibliometric data has not only been applied when researching various aspects of science but also has been firmly established as an integral part of research evaluation and university rankings. Meanwhile, there is a large part of DH that can be found in informal scholarly communications where scholars build connections, establish collaborations and exchange ideas outside of the formal channels. Social media analysis, on the other hand, can provide a ‘material mirror’ that can be used to decipher the complexities of a community and reveal personal connections that cannot be reflected via formal publications (Burton, 2015, p. 5).

As will be discussed below, the two datasets span different time periods, and they complement each other while their overlap of time also provides opportunity for comparison. The bibliometric dataset contains journal articles published between 1966 to 2017, while the Twitter dataset contains tweets posted between 2006 to 2017. There is also space for tracing the field’s knowledge dissemination and publication forecast through dataset analysis and correlation. Data extracted from publications show the knowledge base of a field, and data from Twitter show how the early ideas are raised, and collaborations are built. Although due to citation lag and long process of publication, bibliometric dataset often reflects content that is older than publication, Twitter dataset can supplement this and reveal timely content as soon as scholars posted them. However, scholars on Twitter normally post a variety of different topics and many of them are not DH-related nor academic. The bibliometric dataset, on the contrary, is formally peer-reviewed and contains topics that are specifically related to DH, and thus, can complement the Twitter dataset. In addition, scholarly communities can have different representations across different platforms. As will be discussed, DH communities identified from the two datasets have less overlap than expected, and this not only offers an opportunity for understanding ‘who we are’ from both channels, but also helps to emphasise the diversity and difference in DH communities. By combining and comparing datasets from formal and informal communication channels, this study can better address and discuss the research questions from heterogeneous perspectives.

3.1.1 Data from journals

Bibliometric studies (i.e., a formal channel, such as citation network analysis) can be used to examine and visualise a certain invisible college (Zuccala, 2006, p. 157). This study collects DH journal data for constructing two bibliometric networks in order to study the formal communication channel in DH.

The formal system of scholarly communication can be said to have started in the 17th Century (Meadows, 1980, pp. 1–24). Although the character and structure of scholarly documents have changed significantly (Gross et al., 2002a, pp. 214–228), the key components of this system have remained over the centuries – the bibliometric elements, e.g., the author byline and source records (Koku et al., 2001, p. 1754).

Studies of these key components have, therefore, formed the field currently known as Bibliometrics, which normally starts with the metadata of the basic bibliographic elements, such as author names, article titles, keywords, journal names, and cited references (Rafols and Meyer, 2010). Among all these elements, the practices of the source records (i.e., the references) and author byline (i.e., author names), in particular, have stood out and become not only stylistic choices but also important elements of disciplinary self-identity (Hellqvist, 2009, p. 315). As academic communities grow, individuals cannot know everyone in their field, and thus, there is a need to develop an agreed system of authority. Accordingly, on one hand, the application of references increased significantly from the 17th Century until the current status, with initially only about one third of articles published with references, until almost every research article in the 21st Century (Gross et al., 2002b). The author byline on academic publications, too, has attracted interest from a variety of researches, such as author contribution studies (Mattsson et al., 2011) and author collaboration studies (Ding, 2011; Köseoglu et al., 2018).

As mentioned in chapter 2, given that there is no off-the-shelf DH bibliometric dataset to download, some researchers have constructed datasets by searching keywords in relevant databases (e.g., Leydesdorff and Salah, 2010, p. 787; Zhao and Strotmann, 2011, p. 118), and others chose a journal-based method to collect data from particular journals (e.g., Wang and Inaba, 2009a; Tang et al., 2017).

The method of keyword searching, however, is not practical when studying DH. Firstly, as many different terms were used to refer to this discipline, a set of keywords might not be sufficiently representative. Further, even if there were a set of acceptable keywords, many DH-related articles and publications might not use them as keywords or add them in their article titles. A search of 1,195 *LLC/DSH* articles published during 1986-2017 (*Digital Scholarship in the Humanities*, *DSH*, formerly known as *Literary and Linguistic Computing*, *LLC*) found only 40 titles (3.35%) that include either 'digital humanities' or 'humanities computing'. In addition, although using keywords might be a common publishing practice in most science fields, many DH specific articles, such as publications in the journals *DSH/LLC* and *Digital Humanities Quarterly* (*DHQ*), have not used any keywords, nor do they include any classification terms in their titles. Moreover, studies that do include DH related terms as their keywords or in their titles (e.g., linguistic, or GIS) might not actually be about DH. 'Digital humanities' is not a category of specialty in either Web of Science or Scopus at the time of writing. The most relevant categories in Web of Science are 'information science library science', 'humanities multidisciplinary', or 'computer science interdisciplinary applications'. In Scopus, categories are even more general, such as 'social sciences', 'computer science', and 'Arts and Humanities'. Hence, methods that are based on keyword searching cannot identify a more comprehensive list of DH publications than by simply collecting a particular set of journals.

Collecting publications from specific journals seems more tangible, nevertheless, selecting which journals should be collected is not straight forward. Although some bibliometric studies took Leydesdorff and Salah's quasi-JCR journal citation evaluation (Leydesdorff and Salah, 2010, p. 787) into consideration when selecting journals in other disciplines, their evaluation is based on journal data drawn from the A&HCI (Arts & Humanities Citation Index) and the JCR (Journal Citations Report) in the Web of Science where key DH journals have not been included (when the current study collected data in January 2018). Apart from these databases, at the time of writing, DH journals have not been indexed in Scopus either (e.g., *Digital Humanities Quarterly* was not indexed in these repositories until 2018). Therefore, the quasi-JCR journal selection is not applicable to study DH.

Even though some DH articles are indeed indexed in some common databases, there are restraints in these sources that need to be improved. Their collections are not representative, especially in the humanities and social sciences (Hicks, 2005, pp. 473–474). For example, according to 2017 JCR from Web of Science, there are many more publications in the natural science categories (e.g. Economics, Mathematics, Biochemistry & Molecular Biology) than in the humanities or social science categories (e.g., Psychology, Education & Educational Research, Information Science & Library Science) (Web of Science, 2017). In addition, their metadata lacks support to all-author citation analysis as Web of Science only indexes the first author of each cited reference. Although Scopus does index up to eight authors, it is still not enough when assessing highly collaborative publications and works, which are not unusual in disciplines like DH. Collecting the data of all the authors is crucial when it is used for analysing and evaluating influential scholars who collaborate very often. Digital Humanities is often closely associated with its collaborative nature which makes it stand out from traditional Humanities. Although single-authored papers are still predominant, the co-authored papers have been increasing in DH (Nyhan and Duke-Williams, 2014a, p. 387). Apart from that, different disciplines have different publication cultures, and sometimes the lead scholar (i.e., the head of the research team) is put as the last author of the publication in disciplines such as medical science (Sonnenwald, 2007, p. 670).

This thesis builds a DH bibliometric dataset that covers a period longer than any other existing dataset. It collects all the publications (3,251 articles with 49,047 cited references) of the three most important DH journals within the range from 1966 to 2017 inclusive. This dataset will be published openly and so can also be reused by other DH bibliometric studies providing them with comprehensive and cleaned data and helping them save the time and labour of data collection and cleaning.

In the current study, the three journals listed below were selected as the data sources due to their importance and influence in DH and ADHO. As discussed, journal-based data collection can be more inclusive and convenient to manage compared to other approaches.

Computers and the Humanities (CHum) was arguably the field's first journal. It was established in 1966 and ceased publication in 2004. During that time, it published

many of the latest works on computer applications in the humanities as well as pedagogical practices. Until 2004, it was the official journal of ACH (The Editor, 1993). Although it no longer publishes, it played (and is still playing) an important role in providing a platform that holds crucial works for DH scholars.

Digital Scholarship in the Humanities (DSH), previously known as *Literary and Linguistic Computing (LLC)*, not only plays an essential part in publishing DH works and supporting DH developments, but also in forming DH communities and organising DH events. It is the official journal of the EADH and the ADHO. Its subscription not only links the membership of both associations but also confirms the eligibility to apply for relevant bursaries and prizes.

Digital Humanities Quarterly (DHQ), is one of the official journals of the ACH and the ADHO. Unlike *CHum* and *DSH*, it is fully Open Access and born-digital, and it aims to 'straddle the print/digital divide' (Flanders, 2019). In 2007, it was founded as an experiment that embraces dynamic publication formats, open standards of content, and diverse author groups (Digital Scholarship Group, 2012).

These three journals are understood to be representative of the DH field and have been analysed by most DH bibliometric studies (although there are few). For example, Wang and Inaba applied keyword analysis to articles in *LLC* and *DHQ* (Wang and Inaba, 2009a, p. 123), Bowman selected the same journals for a large-scale ACA research (Bowman et al., 2013, para. 3), and Nyhan and Duke-Williams used data from *CHum* and *LLC* for their co-authorship study (Nyhan and Duke-Williams, 2014a, p. 387).

Given the time and scale of this PhD study, a plan for collecting from other sources was considered impractical. However, there are other journals, books, as well as conference proceedings that distribute significant DH-related works, such as *Digital Studies / Le champ numérique*, (Gold, 2012; Gold and Klein, 2016; Schreibman et al., 2004a, 2016), ADHO annual conferences, and many more. Still, compared to previous DH bibliometric studies that used mostly one or two journals for up to 10 years' data (reviewed in chapter 2 Literature Review), the data range in this study is considerably wider and more comprehensive.

3.1.2 Data from Twitter

Twitter was launched in 2006 and has had a growing number of users across the world. Users on Twitter can post and interact with microblogs that are called 'tweets'. One tweet is restricted to 280 characters (was 140 before November 2017) except for Chinese, Japanese, and Korean (still 140 characters). In 2013, Twitter became one of the top ten most-visited websites and was hailed as 'the SMS of the Internet' (D'Monte, 2009; Alexa, 2013). In 2017, Twitter had more than 330 million active users monthly (Molina, 2019).

There are many ways to make connections with other Twitter users, and the most common ones are 'follow', 'mention' (also known as 'at' or '@'), 'reply', 'like', and 'retweet' (or 'RT'). Users can 'follow' any other users so that they can subscribe and see the tweets of the accounts they follow. Some studies have pointed to the research significance of the number of followers when selecting users for Twitter analysis studies, such as that the subject users should have more than 1,000 followers to be influential enough to study (Côté and Darling, 2018), while others chose to research on the whole scholarly community regardless of the number of followers (Ross et al., 2011; Grandjean, 2016). Users can 'mention' ('at' or '@') other users by tweeting or replying '@' with their user handles. In this way, they can directly address other users, and the addressees will receive notifications when being mentioned. Also, tweets posted on Twitter can be 'liked' and 'retweeted' so that messages can be distributed to a wider audience.

There are other functions, such as 'hashtag' and 'direct message'. The hashtag ('#') is used to mark keywords or topics in a tweet. It was created originally by Twitter users as a way to categorise posts. Many events, such as academic conferences or political movements, use official hashtags to create a tweet stream and enable related messages to be linked and collected in one place. The direct message is when two (or more) users follow each other; they can message each other privately or set up group chat in the same way.

All of these features of Twitter cut down the effort that is required for the users to create, disseminate and digest information. The conveniences formed by Twitter also produce an 'ambient intimacy' that enables people to keep in touch with others with a level of

frequency and intimacy that would not often be possible due to the limitations of time and space (Reichelt, 2007).

In general, scholarly activities on Twitter are meant to be much more public, so it is relatively easier to discover the entire history of a scholar on Twitter by the tweets, retweets, likes and mentions than on other media; there are, however, settings to keep some or all the activities private, as well as to hide or block any other users. Features like these create many connections among users, and these make Twitter an ideal data source to study online social networks and interactions. Twitter analysis benefits from many quantifiable features, such as the off-the-shelf number of tweets, retweets, and likes that are trivial to collect (Eysenbach, 2011; Peoples et al., 2016; Thelwall et al., 2013). With its openness to API (Application Programming Interface), it is relatively convenient to get clean and structured data about these connections without needing significant data cleaning. Because of these advantages, Twitter has attracted academic attention from 2007 (Java et al., 2007), and later became one of the mainstream subjects in social media studies (Williams et al., 2013, p. 385).

As the most popular microblogging platform in the West, Twitter not only contributes to the formation of the DH community as a digital 'backchannel' during academic events (Ross et al., 2011), but also provides a social network for DH scholars to communicate and exchange ideas while keeping all the tweets as informal interactions in one place (Quan-Haase et al., 2015a). Although not all DH scholars have used social media, for many of them, it has become a necessary part of their academic practice. Learning to tweet has even been perceived by some DH scholars as a key way of academic communication (Quan-Haase et al., 2015b, p. 3).

As some have argued that it is becoming increasingly difficult to see the networks of this 'expansive, movable, but precarious' field of DH under its 'still not big enough in terms of diversity and access' tent through narratives or retrospectives (PMLA Editorial Board, 2018), it is critical to understand how and why Twitter is being used and what are its implications in DH.

This PhD study selects scholars (or users) from authoritative social accounts and filters data according to others' recognition. In total, this study collected 3,154 users

and six million tweets published from 2006 to 2017. The process of data collection and analysis is demonstrated in chapter 5, DH Twitter Network Analysis.

Given the time and scale of this PhD study, collecting data from other social media sources was considered impractical. However, apart from Twitter, online discussion groups such as *fora* are also useful for providing access to the academic social world (Matzat, 2004). Before Twitter was widely used, DH scholars had (and still have) many discussions about DH intellectual, scholarly, pedagogical and social issues on the online discussion forum – the *Humanist* (Nyhan, 2016). Additionally, research shows that traditional face-to-face communication is still one of the many preferred communication methods for scholars (Koku et al., 2001). Video chat applications like Skype (Kay and Lauricella, 2015, p. 5) and Google Hangouts (Chan et al., 2015, p. 171) are studied and used for academic communication as they become increasingly popular nowadays (Jones et al., 2016, p. 1124).

3.2 Subject Specialty

The *subject specialty* of an invisible college can be studied via a collection of academic publications that carry the research practices, culture, values and rules of this invisible college (Sandstrom, 1998, pp. i–iii; Zuccala, 2004, pp. 1–2) as well as via social media data and, in this study, the Twitter hashtag (Doctor, 2012; Türker and Sulak, 2017). Based on publications and Twitter hashtags, the *subject specialty* can reveal the intellectual structures of a field through two different aspects (Zuccala, 2006, p. 156).

3.2.1 Author co-citation network¹⁶

The continuous use of citation in academic writing, in particular, established a rational convention across disciplines, and citation was referred to as one of the important

¹⁶ The terminology should be noted in this section. The terms used in citation analysis can be very easily misrepresented with the words people use in daily life. It is common for the term ‘citation’ to be used interchangeably for either ‘reference’ or ‘bibliography’ with different contexts providing the meaning. Similarly, the concept of how authors make references is called either ‘citing behaviour’ or ‘referencing behaviour’. When article A makes a reference to article B, it is often said that A cites or references B, and B is cited by, receives a citation from, or is one of the cited references in A. In this study, it is the references made in collected articles that build up the dataset, and not the later citations to those published articles, although both words will be used when necessary.

indicators of contributions to knowledge (Cronin and Sugimoto, 2014, p. 4). The modern practice of examining scholarly production in this way began with Garfield's research on the Science Citation Index (Garfield, 1955), and he raised the ideas of metrics and impact factor that led to other measures such as the *h*-index (Hirsch, 2005) and the construction and operation of bibliographic databases (e.g. Scopus and Web of Science).

Even though there are many criticisms, by using citation analysis methods, one can analyse the discipline through its unique referencing culture and the ways scholars use and cite references (Hellqvist, 2009, p. 316). Ever since Garfield demonstrated that bibliometric analysis was an efficient evaluative tool of academic production, although many contest the validity and usefulness (Greenseid, 2008; Moed, 2006), citation has become the central interest and main focus to understand how scholars communicate formally and acknowledge others' contributions via publications (E. Garfield, 1979).

Since measures such as impact factor have raised many negative debates in information science and other disciplines (Seglen, 1997, p. 497), this study will only concentrate on the network visualisation instead of evaluating or assessing the scholars by the number their citations. Although some weighting methods are applied, this study chooses not to interpret these results alone as indicators of academic productivity.

Citation network analysis, as the first step of the Invisible College model, is one of the broadly used bibliometric approaches that visualises the academic networks of authors, documents, journals, institutions, etc., based on their quantified relationships drawn from the citation data (Usdiken and Pasadeos, 1995, pp. 503–505). It has been applied by many studies to explore invisible colleges of a variety of disciplines, such as General Science (Wagner, 2008, pp. 5–6), Economic Analysis (Verspagen and Werker, 2003, p. 395), International Law (Hernandez, 2016), Mental Disorder Study (Blashfield and Reynolds, 2012), Public Administration (Algarni, 2014), Accounting (Casanueva and Larrinaga, 2013), Entrepreneurship (Teixeira, 2011), and Journalism (Chang and Tai, 2005).

Author co-citation network analysis, or Author Co-citation Analysis (ACA), is one of the citation network methods that has assisted many studies (Eom, 2003; Gmür, 2003; White and Griffith, 1981) to discover their field structures, trace knowledge distribution, explore resources, investigate research impact, and explore scholarly communication. Especially with current advanced computational techniques, many new developments in ACA have made large-scale projects achievable (Backhaus et al., 2011; Ravallion and Wagstaff, 2011; Zhao and Strotmann, 2015), even the network visualisation with data from more than 1,000 journals (van Eck and Waltman, 2010, p. 537).

Because of these advantages, ACA is particularly useful and suitable for answering the DH *subject* question and mapping the DH intellectual structure. As discussed in chapter 2 (Literature Review), in recent years, more and more DH scholars have applied bibliometric analysis and data visualisation methods to DH publications. However, no study has applied ACA to investigate the DH intellectual structure due to the lack of a comprehensive bibliometric dataset such as the one this study has built.

ACA is seen as a 'rear-view mirror' that can help to reveal the knowledge map of a field, and this is particularly what DH needs – to provide an objective field image to assist in its essential and ongoing disciplinary debates¹⁷. More specifically, it selects authors as the nodes to display on the network and calculates the total times that two authors are cited together by a third party, and that number indicates the 'research distances' (i.e., edge) between each of the two authors within the discipline. In other words, when an article cites at least one paper of author A and at least one of author B that is different from the one of A (in any author position), the co-citation count of authors A and B increases by 1. Thus, ACA reflects that the more the two authors are

¹⁷ Another similar citation analysis method that helps to discover the relationship between authors and subjects is Bibliometric Coupling Analysis (BCA). Contrary to ACA, calculating the BCA value between two authors is defined as when author A and author B are writing two separate articles, and they both cite the same paper, then their BCA value increases by 1. The more articles they both cite in their list of references, the closer their researches are connected. Compared to ACA, BCA investigates more about the knowledge front of a field and has shorter lag time, while ACA discovers the intellectual foundation and knowledge base. Consequently, BCA is more suitable for discovering new subjects and knowledge of a field while ACA is more suitable for tracing the history and knowledge formation of a field. Because of such difference, this study chooses to use ACA to better answer the *subject* research question and help uncover the DH 'hidden' history. This choice has also been supported in many other disciplinary studies for its practicality (Zhao and Strotmann, 2015, p. 38).

cited together by other articles, the closer the two authors' research topics are connected even though they did not cite each other directly (Griffith, 1989).

The ACA research procedure usually includes: citation data collection; data cleaning and name disambiguation; the calculation of author weights; the co-citation count; the accumulation of the author co-citation matrix; the network visualisation of the whole period (McCain, 1990).

Firstly, to clean the dataset and identify the authors, this study applies an all-author approach, i.e., all authors were counted no matter what their position on the author byline for the published article. As discussed above, although the first author in many publications represents the most significant contribution, other orderings are common in specific fields. For instance, the last author in the byline may be the team leader whose contribution is also important, and alphabetical ordering is also common (Sonnenwald, 2007, p. 643). As DH is believed to be an interdisciplinary field where different byline conventions might be found, the all-author approach can help better to analyse the field inclusively. This study thus treats each author equally no matter what their position in the byline.

Secondly, although this study does not focus on quantifying the author contribution and does not seek to enter the area of evaluative bibliometric analysis to rank and evaluate the research quality or productivity of authors, weighting measures are still needed for visualisation and indication, which should also be discussed. Fractional citation count is believed to be more helpful when dealing with calculation bias compared to a full citation count (Perianes-Rodriguez et al., 2016). It distributes the number of citations according to the number of authors listed on the cited publication (i.e., weighting each citation as $1/n$, where n is the total number of authors in the cited publication), and it is an efficient method to justify differences in scholarly citation counts among different scholars from different backgrounds (E. Garfield, 1979; Moed, 2010). Although some note their disagreement (Radicchi and Castellano, 2012), fractional count has been preferred by many scholars for its emphasis on collaborative works (White and McCain, 1998, p. 327; Zhao and Strotmann, 2015, p. 28), and it could ease the citation gap between the 'newcomers' and the 'elders' (Leydesdorff and Ophof, 2010, p. 2367). This study, thus, chooses the fractional method (i.e., *fractional all-author citation count* to weight DH scholars, and *exclusive all-author co-*

citation count to calculate the co-citation matrix) for its fitness to handle the DH dataset – a large amount of citation calculation of an interdisciplinary, collaborative and changing field (Ahlgren et al., 2003; Lindsey, 1980; Zhao and Strotmann, 2011). A more comprehensive process of the calculation will be presented in chapter 4 (DH Bibliometric Network Analysis).

After the weighting and calculation of the author co-citation matrix, the data visualisation can be produced by various software packages (e.g., Gephi, SPSS, R, Sci2 Tool, Pajek, VOSviewer). Every node on the network represents a cited scholar, and the distance (i.e., edges) of these nodes describes the relationships of authors by other citing articles according to the citation data (White, 1990).

When applying ACA, the citation lag time should also be taken into consideration as a common point. In order to accumulate enough citations for an article and to construct a citation dataset for co-citation, it needs a lag time usually from five to eight years (Hopcroft et al., 2004, p. 5250). Given the journal publication process (e.g., a significant time for peer-review, revision and typesetting), it could be even longer for an idea to be acknowledged and cited by other readers. Yet, even with this citation lag time, ACA in turn, provides a favourable opportunity to study the earlier achievements of the discipline, the knowledge base of a field and to trace back to its disciplinary origins.

3.2.2 Twitter hashtag co-occurrence network

While the ACA network shows a ‘rear-view mirror’ of the knowledge structure based on formal publications, there is a large part of DH that can be found in informal scholarly communications where scholars build connections, establish collaborations and exchange ideas outside of the formal channels. With the help of data from Twitter, more information about DH subjects can be explored.

Given the dynamic nature of Twitter (Hogan and Quan-Haase, 2010), it is not surprising to find that scholars in different fields have been using it differently (Archambault and Grudin, 2012). For example, scholars in DH tend to use Twitter more often than those in Biochemistry and Economics (Holmberg and Thelwall, 2014). However, previously little is known about why DH scholars use it so often and what topics they have been discussing on Twitter.

As above, this study aims to answer such questions by analysing the Twitter data, but Twitter studies can be conducted in various ways based on various types of Twitter datasets. It is necessary to select the most effective method. Williams et al. grouped existing Twitter studies into four overlapping categories according to their subjects – *message*, *user*, *technology* and *concept* (Williams et al., 2013, p. 394). The majority were about *message* (text and metadata) and *user* (identities and connections) studies while *technology* and *concept* have drawn less attention, which is understandable as the latter is more about the introduction of new features or interfaces, general reviews, or overviews. Given the distinct purposes of the four types, this study chose to conduct *message* study to investigate the DH Twitter subject structure and apply *user* study to examine the DH social network (which will be discussed in the next section 3.3.2 Twitter co-retweet network).

Among all the *message* studies, the hashtag in particular has attracted the greatest interest, and it is one of the most popular narrative forms of current digital activism (Yang, 2016).

The tagging practice is not only considered to be a way to summarise the content of a message, as a socially defined taxonomy, but also plays a reflexive role for commenting on a given topic (e.g., #transformDH), and is described as ‘a node of continued context’ across conversations ‘between what is contextual and what is chronological’ (Rambukanna, 2015, para. 4; Eriksson Krutrok and Lindgren, 2018, p. 3). In this way, hashtags allow users to extend their discussions outside of their normal social networks linking different user groups and times with the same topic.

Before hashtag analysis became popular, many Twitter *message* studies had challenges examining Twitter contents. Given that most tweets are original, coded, unstructured and contain many abbreviations, it is often difficult and inefficient to carry out content analysis on Twitter *message* using traditional methods such as text analysis. This is due to the tweet length limitation (i.e., 140 characters before November 2017), and users have invented their own ways to solve this problem, such as new shorthand, codes and jargons. These solutions have made tweet text analysis more difficult to clean and find patterns. Many studies have highlighted the need for new ways other than traditional text analysis to explore the Twitter data (Ross et al.,

2011, p. 223), and the hashtag, therefore, has gradually become the favourite in Twitter content analysis (Carrotte et al., 2017; Ogan and Varol, 2017)

Studying hashtags not only takes advantage of the structural feature that Twitter data offers, but also has the ability to demonstrate information dissemination beyond the conventional ‘follower-following’ network (Eriksson Krutrok and Lindgren, 2018). By introducing the method from social network analysis (SNA), this study, therefore, applies the hashtag co-occurrence network to study the DH knowledge structure based on the informal world of Twitter. Hashtag co-occurrence network, as the name implies, counts the number of hashtag co-occurrences in the same tweet, in order to determine whether two hashtags are semantically related (e.g., if they have co-occurrences greater than a cut-off threshold) (Türker and Sulak, 2017). This has the ability to assist in discovering the DH Twitter subjects and further uncover what and how DH themes are formed on Twitter. The network construction procedure is similar to that of ACA and will be introduced in 5.2 (Hashtag co-occurrence network).

3.3 Social Actors

Compared to disciplinary subject studies, *social actors* can provide another overview that was previously unavailable to scholars. With the social technology development, sociometric studies provide a ‘material mirror’ that can be used to decipher the complexities of a community and reveal personal connections that cannot be reflected via subject networks (Burton, 2015, p. 5).

The *social actors* phase investigates the social communications within the invisible college, i.e., the personal communications among predominant scholars, whether formal or informal. Sociometric methods are usually applied to study such questions as, e.g., co-authorship network, or users’ connections and interactions on the social media (Gruzd et al., 2012; Holmberg and Thelwall, 2014; Neal, 2012; Veletsianos, 2012). Data collected from publications and Twitter, as well as other sources (Rowlands et al., 2011, p. 184; Gruzd et al., 2012, p. 2341; Algarni, 2014), are usually analysed and visualised as social networks to explore the ‘multifaceted phenomenon’ of an invisible college (Zuccala, 2006, p. 159).

This section provides a detailed discussion about methods and rationale to construct two scholarly personal networks – co-authorship network and Twitter co-retweet network.

3.3.1 Co-authorship network

Many studies of scholarly collaboration have used the co-authorship method to assist their analysis (e.g., Cronin, 2005; Fagan et al., 2018; Köseoglu et al., 2018), and DH is no exception (e.g., Nyhan and Duke-Williams, 2014a; Tang et al., 2017). Lundberg pointed out that 'analysing co-authored publications has become the standard way of measuring research collaborations' (Lundberg et al., 2006, p. 575). Kumar explains the relationship as:

[...] researchers mostly choose with whom they would like to do research and then pen down the results in the form of a co-authored research paper or artefact. These collaborations leave digital footprints in the form of bibliography, which can be effectively tracked and evaluated. (Kumar, 2015, p. 57)

Co-authorship is an important system that connects different types of specialties to produce research outputs (Ponomariov and Boardman, 2016, p. 1939). It has been 'operationalised' as a proxy for research collaboration for decades by academic evaluators and policy makers not only because the bibliometric data is structured and available but also because it indeed reflects an aspect of research collaboration (Melin and Persson, 1996), although some disagree (Kumar, 2018)¹⁸.

There are different approaches to study co-authorship patterns, such as counting bilateral and multilateral co-authored papers, unique affiliations and regions (Adams and Gurney, 2018), and the percentage of multi-authored papers (Hudson, 1996). Most of the co-authorship studies in DH, as reviewed in section 2.2.1 (Literature

¹⁸ It should be noted that using co-authorship based indicators has been suggested by many bibliometric studies as but one index of collaboration (Ponds, 2009; Gazni et al., 2012; Nyhan and Duke-Williams, 2014a). Uncritical use of the co-authorship method might lead to misinterpretation of scholarly collaboration activities and thus provide incorrect data for decision-making (Lundberg et al., 2006). Therefore, such measures should be handled with care and used as one source of evidence on exploring the scholarly communication and collaboration.

Review, Co-authorship) (e.g., Spiro, 2009; Nyhan and Duke-Williams, 2014a), applied these approaches.

Studying co-authorship from the aspect of social networks, on the other hand, is a relatively young approach. It helped the field of co-authorship studies to reattract interest since Newman firstly applied social network analysis (SNA) methods to explore both the micro and macro patterns of large co-authorship networks (Newman, 2001). After that, co-authorship network analysis has been practiced more widely to study the research communities in many fields and disciplines.

Applying SNA in a co-authorship study helps us to uncover its structure, patterns of connection, and formation mechanism (Kumar, 2015, p. 55). The scholars are the nodes on the network while the co-authorship relation is the edge, which is formed by two scholars who co-authored an article together. The more scholars that co-author publications, the more nodes and edges, and the larger the network will be. Such a network provides an extensively documented view of an academic community and reveals its many structural character aspects, e.g., level of connectivity and field development (macro aspect), central or peripheral scholars in that network (micro aspect).

In DH, there are few studies that have applied social network analysis to visualise the co-authorship structure, apart from the attempt by De la Cruz et al., with a limited dataset (i.e., 178 *DHQ* articles) (De la Cruz et al., 2015). One reason is the difficulty to collect, clean and construct a bibliometric dataset. As mentioned, constructing a well-cleaned and comprehensive bibliometric dataset of DH publications is one of the main contributions of this study, and the process and difficulties will be thoroughly explained in section 4.1 (Data collection and cleaning).

This study applies co-authorship network analysis as the first method to explore the social structure of the DH community. The approach is similar to that of Author Co-citation Analysis (ACA). Compared to ACA (discussed in section 3.2.1) which also applies SNA to examine bibliometric data, co-authorship network focuses more on investigating personal connections and social interactions, while ACA aims to study the knowledge structure and intellectual formation of the field. Co-authorship network studies the personal relations that are formed by both co-authors (i.e., bilateral link),

indicating the two-way interpersonal communication and social connection they have had; citation, however, is often made as a one-way relationship (i.e., unilateral link) from the article to the cited publication from which they carry the influence (or knowledge flow) from the source publication to the citing article. In this way, the co-authorship network not only explores relatively newer author data and the knowledge front, but also focuses more on the scholarly communication and collaboration.

As mentioned in co-citation network section (3.2.1), different counting approaches can be used to weight the nodes (e.g., full counting, fractional counting). Although ACA will use the fractional counting to calculate the node weight, this study chooses the full counting (i.e., an author's total number of publications) to construct the co-authorship network.

As mentioned, one of the main purposes of constructing a network of co-authorship is to study more of the scholarly community in general, instead of evaluating individuals. Relationships and social structures are often more essential than measuring individual authorship. Consequently, many co-authorship studies applied full counting to weigh the nodes. As mentioned in the Literature Review (2.2.1 Co-authorship), previous studies have used full counting methods to weigh the co-authorship in DH. Even though the fractional counting approach is preferred by some bibliometric studies and will be applied in the ACA section of this study (section 4.2), the full counting approach pulls co-authorship clusters closer on the network so the important nodes and bridges are clearer to detect (Perianes-Rodriguez et al., 2016, p. 1186). In addition, the full counting method is easier to explain with integer weighting, and it is especially useful when presenting the results on visual graphs like networks.

3.3.2 Twitter co-retweet network

Twitter has many features that can be used to construct social networks, such as *mention*, *follow*, *like*, *retweet*, and *hashtag*, as discussed in 3.2.2 (Twitter hashtag co-occurrence network). It is important to distinguish the most suitable feature and identify the most appropriate method to address the *social actors* research question.

Firstly, *hashtag* is studied mostly for content analysis purpose to indicate discussion topics and subjects, and it is not suitable for demonstrating personal connections or an individual's influence. *Mention* is researched mostly for its conversational purpose.

Although it is useful to discover conversational threads and hot topics related to a particular time period (i.e., the tweet lifespan), it is not the best choice to help visualise a scholarly community (Guille and Favre, 2015).

Follow seems to be the most direct and apparent indication of social relationship on Twitter. Although it can reflect the subscription choices made by users, studies have shown that a network construct based on the follow-following relationship is more of an informational network rather than a social network (Myers et al., 2014). In other words, while some follows are built on social ties, such a feature is primarily about information consumption. In addition, harvesting *follow* relationships to construct the DH community social network has already been done and thoroughly studied by Grandjean (Grandjean, 2013a, 2014, 2015, 2016). It is more research-worthy to analyse the DH community with new approaches and from other perspectives. This study aims to apply different methods to compile the Twitter dataset and construct the network, and it can thus complement previous studies as well as uncover new results.

Like and *retweet* are the other two very popular features that have been studied by many (Boyd et al., 2010; Suh et al., 2010; Cole, 2015; Giachanou and Crestani, 2016). They are both indicators of tweets and users that have a certain degree of influence. The fundamental difference between *like* and *retweet* is that *like* means that the user enjoys the tweet; users can also *favourite* the tweet which keeps it as a private list that can only be seen by the users. *Retweet*, on the other hand, means sharing the tweet (either with comments or not, and whether likes or not) with all of the user's followers, who in turn might share it with their followers. A *like* is a form of bookmarking, and users can go back and review a list of their likes, while *retweet* is an indicator of greater influence and more extensive distribution that promotes both the tweeter and the retweeted content. This study, therefore, chooses *retweet* as the research subject to study the scholarly social connections and interests in DH.

All the above-mentioned Twitter features can form explicit links (or direct links, where relations are based on direct interactions), as well as implicit links (or co-occurrence links, where relations are built by indirect but common third-party activities) (Kwak et al., 2010, p. 591; Wang et al., 2014, p. 2). Both the two types of links can form a visualised network, but the latter (i.e., co-occurrence link) is believed to be able to demonstrate more information and aspects (Song et al., 2016, p. 10). More specifically,

the co-occurrence link is an indirect link between each of two nodes that are often built upon direct connections, such as: co-mention, co-follow, co-like, co-retweet, and co-hashtag. These co-links are implicit connections between any pair of users.

Although there are some popular network studies that have employed direct-link calculations (e.g., Golley and John, 2015; Grandjean, 2016), and a few researchers think that co-occurrence links may probably bring information loss because of the lack of link directions (Wang et al., 2014, p. 20), more and more network studies prefer co-occurrence links over the direct-link (Davis et al., 1979; Türker and Sulak, 2017; Eriksson Krutrök and Lindgren, 2018). This might be because the direct-link is more about the relationships created by the users' own Twitter activities, while co-occurrence link could bring more knowledge about the relationship of user A and user B in 'other people's eyes'. Some studies proposed to transform the whole system of direct-link methods to co-occurrence methods, in order to show network connections more thoroughly and from more perspectives (Zhou et al., 2005; Wang et al., 2014, pp. 18–19).

This study, therefore, chooses to apply the co-retweet method to build a Twitter social network of the DH community. This method in particular not only helps to indicate the DH social connections that are based on DH scholars' own retweeting activities, but also reveals the similar retweeting interests that users share but have not yet been discovered through direct-links.

3.4 Information Use Environment (IUE)

Studies associated with an invisible college often lead to questions which prompt the need to investigate researchers' backgrounds, such as institutions or working spaces, as well as their personal backgrounds, e.g., country, language, gender, and race (Tuire and Erno, 2001). This is the last stage of the model – *information use environment (IUE)*.

IUE studies can provide additional background and context analysis that helps to interpret the bibliometric and social networks constructed in section 3.3 (Social Actors). It focuses on the voluntary activity of scholars which is defined as the services that scholars perform or the behaviour they conduct to support their research system or to develop their community. For instance, scholars from certain areas tend to engage

more with their local research topics or their native languages (Chen and Hsueh, 2013; Tello, 2016; Mäkelä and Tolonen, 2018). These behaviours include joining institutions, publishing works, collaborating with colleagues, reviewing publications, organising conferences, arranging committees, and participating in building educational programmes (Zuccala and van den Besselaar, 2009, p. 120).

This PhD study explores the background of selected DH scholars by three factors – gender, language and affiliated country. The study assigns gender (mainly by first name) and country information to 3,382 DH scholars (based on co-authorship network) and 3,154 DH Twitter users (based on co-retweet network). These also help to compare and verify the outputs of the bibliometric and sociometric results in this PhD research. Only the public information was collected and there will be no link from the dataset or results to identify any private information¹⁹. The ethical application for this study has been approved by the appropriate departmental tutor, and there is no further ethical approval or GDPR application needed. We are aware that apart from gender, language and country, other forms of data sources, e.g., race, ethnicity, sexual orientation, socio-economic status, age, physical abilities, religious beliefs, political beliefs, are also important determinants. These factors show impact that can be investigated further in the future.

3.4.1 Gender

Although age and academic position usually have more significant impact on academic production and communication, other variables, such as gender, also play an important part (Rørstad and Aksnes, 2015, p. 329). Studies have shown that the difference in publication rate between male and female scholars is significant in many fields (e.g., natural sciences, technology, medicine) (Noordenbos, 1992; Aksnes et al., 2011). Based on data collected in Norway, a country considered to be one of the most gender equal countries in the world, in the natural sciences, gender accounts for 22% of the difference in publication rate in favour of men, when all other variables are

¹⁹ The dataset of this background information is held in a local machine and protected by encrypted password according to the College's data protection guidelines.

constant, and the number is 15% in engineering and technology and 8% in the medical field (Rørstad and Aksnes, 2015, p. 327).

Gender difference is not only significant in the publication rate, but also in scholarly communication and networking. Different studies have shown that gender has a significant effect on the network outcomes (Brass, 1985; Ibarra, 1992), with female groups mostly having either a significantly higher centrality (Brass, 1985) or significantly lower centrality (Tharenou, 1999) than male groups. Despite some results showing low contribution rates of female scholars in collaborative activities (Kretschmer and Agillo, 2005), more recent works show otherwise. They argue that female scholars are more likely to be in and benefit from central positions within the co-authorship network because of their stronger and confident characteristics that helped them to break the gender publication barriers (Badar et al., 2013).

Why does gender result in such significant differences? Some think that female scholars tend to publish fewer publications because there are fewer female scholars climbing the academic ladder (Long, 1992; Kyvik and Teigen, 1996; Lawrence, 2006; Abramo et al., 2009). Usually, the proportion of female researchers lessens along the academic hierarchy, starting with a balanced gender rate for PhD students but later with predominantly male scholars in senior positions, such as professors (National Science Foundation, 2017). Others explain this difference by showing that women and men choose differently. They argue that female scholars allocate more of their time on teaching and administrative tasks, while their male counterparts focus more on research and projects (Rørstad and Aksnes, 2015, p. 318). Nevertheless, this is not always the case. Another study found that young female researchers outperformed young male researchers in research and the number of publications (van Arensbergen et al., 2012).

Others found that although, as the age increases, there are fewer female scholars, their publication rate grows continuously by age especially after 59 years old (Rørstad and Aksnes, 2015, p. 318). When scholars reach the age of 70, mainly professors or retired professors, the publication of female scholars (1.6) is much higher than the male counterpart (1). This also indicates that female scholars tend to be more active in research even in advanced age.

Gender is particularly important in a field like DH that is predominately male-oriented in terms of the number of publications (Risam, 2015a). More recently, feminist studies in DH have attracted increasing attention, partly due to the increasing criticisms of representativeness and globalisation in DH (Liu, 2012b). Some have pointed to the exclusion of female scholars in the development of DH, and acknowledged their contributions (e.g., Nyhan and Terras, 2017; Wernimont, 2018). Others, such as Wernimont, focus on how to use technology to enhance the discussion about feminism and inclusivity (Wernimont, 2013). Some DH scholars have been directly addressing feminism and women in DH (e.g., Posner, 2015; Nowviskie, 2015, 2016b), and conducting relevant projects, e.g., #transformDH (Bailey et al., 2016), Black Girls Code (Bryant, 2011) and Women Writers Project (Northeastern University, 1986).

However, little is known about what the exact gender difference is in DH in terms of publications and active social interactions. By adding gender information to the co-authorship and co-retweet networks derived from the previous two stages, this study provides a more comprehensive view of the gender distribution in DH co-author and social media communities.

Assigning gender is part of the data collection procedure that mainly includes extracting the gender information of authors as well as Twitter users. In the current dataset, there are 3,382 DH scholars (based on co-authorship network) and 3,154 DH Twitter users (based on co-retweet network). Ideally, the analysis would be more accurate if authors could provide their own gender and country information. A few research projects did use author-provided data for their analysis (e.g., Goswami et al., 2009, p. 214). However, it was not feasible to contact this number of authors and ask for their data within the time period of this PhD study.

In order to identify gender in such a large cohort of authors, some studies proposed automatic approaches based on the author's use of language. For example, Rangel et al., analysed the stylistic features in authors' English and Spanish written texts from social media and the frequency of their use of different grammatical categories (e.g., pronouns and verbs) to obtain their gender and age (Rangel and Rosso, 2013, p. 177; Rangel et al., 2018). These automatic gender identification approaches have attracted much academic attention for their potential in online forensic, security, and marketing studies to help with identifying fake profiles and the senders of harassing messages

across different languages (Fatima et al., 2017). Nevertheless, they are not suitable for identifying author genders in the current dataset for the lack of sufficient texts (e.g., only article titles, abstracts and the profiles of Twitter users are available), and certainly not applicable to articles that have multiple authors from different age and gender groups.

Some studies in DH have relied on ‘gender guessing’ to assign gender which is based on a combination of hand-coding and automated inference (Weingart et al., 2016, pp. 2000–2016). This current study takes such an effort a step further and applies a well-tested name-gender assignment method as proposed in (Larivière et al., 2013; Sugimoto et al., 2015). Three gender categories were created for assignment: ‘female’, ‘male’, and ‘unknown’. It is noted that some people are gender diverse, but the sources for that information are very limited, so this study follows the previous gender category convention (Rørstad and Aksnes, 2015, p. 321). Firstly, a given name list with gender information was developed based on the universal and country-specific name lists in Larivière and Sugimoto’s studies. The country-specific names were applicable to authors from English-, French-, Korean-, Lithuanian-, Persian-, Portuguese-, Serbian-, Ukrainian-, Thai- and Japanese-speaking countries, as well as authors in India. Chinese names, in addition, were mostly assigned by searching the Internet and checking personal pages as the PhD candidate is a native Mandarin speaker. Besides this, author names from other regions employed the universal name list for gender assignment. If it is a unisex name, and there is no additional author information that can be found from the Internet, the author is assigned to the ‘unknown’ category.

3.4.2 Language and affiliated country

When being in a group, people’s behaviour can be influenced by certain social forces in which the group was classified, and this social group has its own perceived environment that is called the ‘definition of the situation’ (Lewin, 1936). When authors are from different institutions, collaboration ties among the ones from different locations and who speak different languages are found to be weak, while others have been found stronger where they are in the same locations and speak the same languages (Tuire and Erno, 2001, p. 494). Factors such as language and country (and sometimes, culture) are believed to be deeply intertwined with each other, and it is not

practical to study them separately when analysing the formation of a community (Goodman, 2011).

As a particularly language-oriented field (or at least for some DH topics), language and country are believed to be critical determinants of the DH community formation (Flanders, 2016; Pitman and Taylor, 2017; Tello, 2017). Moreover, the longer scholars progress up the academic ladder, the more they accumulate influences (Rørstad and Aksnes, 2015). The Matthew effect, i.e., 'the rich get richer and the poor get poorer' (Gladwell, 2008), is a vivid example of the social dynamics that such factors could cause. This concept is not only applicable to the formation of society, but also to the academic community, such as the 'cumulative advantages of academic capital' (Merton, 1968).

In general, many agree that English is the dominant global medium of scholarly publication. In 2004, 74% in Ulrich's Periodical Directory and more than 90% of the social science articles in the Institute for Scientific Information were published in English (Lillis and Curry, 2006, pp. 3–4). Studies have shown that in different non-English-speaking countries, e.g., Poland (Duszak and Lewkowicz, 2008), Portugal (Bennett, 2011), Iceland (Ingvarsdóttir and Arnbjörnsdóttir, 2013), Spain (Moreno et al., 2012), and many others (Flowerdew, 1999), there is an increasing pressure to publish academic works in English. What is more problematic is that English publications are often assumed to be of higher status than those in other languages. Many scholars seem to accept that this English premium has influences that are decisive and important to their opportunities for promotion and research grants (Flowerdew, 2000; Mahony and Gao, 2018).

Some scholars are concerned that writing in English for non-Anglophone scholars not only creates barriers for them to disseminate their work, but also poses challenges and limits participations when they communicate with other scholars (Uzuner, 2008). Some even consider that non-English speakers are 'linguistically disadvantaged' compared to English-speaking scholars when it comes to publishing in international journals (Ferguson et al., 2011, p. 45). There are many studies criticising this 'inequality' that may often lead to discrimination and isolation in the global academic environment (Ammon, 2012).

However, such a situation also has a positive side. Whether scholars' first-language is English or not, the majority would benefit from publishing works in English. By doing so, they gain both visibility in the international academic community and recognition in their own regional society (Bocanegra-Valle, 2014, pp. 65–66). Because of this, we can see an increasingly high flow of submissions to English journals, and many of these have significant influences in their disciplines (Bocanegra-Valle, 2014, p. 67).

Due to the strong influences of English language publications, much attention has focused on the Anglo-American areas, and seemingly, many influential scholars are working in such areas. More and more studies have raised the questions of extending the meaning of 'international' beyond just the Anglo-American countries (Paasi, 2005; Earhart, 2018). This movement indicates the change and development towards a more diverse and inclusive global academic environment, and DH also aims to contribute to such a movement (Mahony, 2018). As text is still the most popular subject in DH (Siemens, 2016; Weingart and Eichmann-Kalwara, 2017), language and country play an important part in most DH research collaborations (Flanders, 2016).

This study, therefore, collects the author and country information by scholars' affiliations provided on their publications as well as on their Twitter profiles. As discussed in section 3.4.1 (Gender), asking scholars to provide their own country and language information is not possible, and an automatic identification system is not accurate. This study applies a similar data collection procedure here to that of the gender study.

More specifically, a list of countries and regions is used based on the information from the United Nations Member States webpages.²⁰ If the author has affiliations from more than one country, the most used country is selected. For example, Willard McCarty lists affiliations in both the UK and Australia, and this study selects the UK as his affiliated country as he published most articles with the UK affiliation. If two or more countries are used with equal frequency in the affiliations, the selection is then made based on personal knowledge and a web search (e.g., the most recent affiliation is selected). Where a country no longer exists (or is now identified by a different name),

²⁰ The country and region list from <http://www.un.org/en/member-states/> retrieved on 2017-10-16.

such as the Soviet Union, as noted for example in an LLC article (Tambovtsev, 1987), then the name of the current country of the institution is used (in this case, the Russian Federation). If no country information could be found, then the author is assigned to the ‘unknown’ group.

3.5 Summary

To sum up, the refined invisible college research model provides a cohesive and comprehensive framework to study the DH community. This research model connects different research aims and methods from both quantitative and qualitative perspectives and combines them into an ‘all-in-one’ model that explores the DH *subject, scholar and environment*. This model not only inherits the advantages of individual methods but also systematically integrates procedures so that each method complements the others.

To the best of the candidate’s knowledge, this study is the first to apply four network methods to visualise and compare the DH subject and community. The model developed in this study can also be applied in many other disciplines exploring their subjects and scholarly communities. It is able to deal with large scale datasets and can be used repeatedly during different disciplinary development stages.

4 DH Bibliometric Network Analysis

To better formulate and narrow down the research questions of this thesis within the bibliometric framework, this chapter aims to answer the following questions:

- a) *Subject*: What are the main topics (or subject specialties) in DH publications? How do they relate to each other? How have they developed over time?
- b) *Scholar*: How collaborative are scholars according to DH publications? What social structure can be identified from co-authorship patterns? What might be the determining factors of the co-authorship relations?
- c) *Environment*: How diverse are the backgrounds of DH scholars (i.e., gender, affiliated country) based on publications? How do gender and country diversities intersect the two above (i.e., the intellectual structure and scholarly communication)?

This chapter outlines the series of detailed steps that were taken to construct two bibliometric networks: the author co-citation analysis (ACA) network and the co-authorship network. The former was constructed based on the co-citation relationships of cited authors while the latter was constructed by the co-authorships of publishing authors.²¹

The first section introduces a compilation of the DH bibliometric dataset that was used for both network analyses (4.1 Data collection and cleaning). Publication metadata was extracted from the three most important DH journals published over 52 years (1966-2017). Following this, the two networks' construction procedures, results, and interpretations are demonstrated separately in section 4.2 (ACA network) and 4.3 (Co-authorship network).

²¹ When discussing network studies, terms such as 'node', 'actor', and 'author' have been used interchangeably, and terms like 'edge', 'link', 'relationship', 'co-citation' (for ACA study), and 'co-authorship' (for co-authorship study) have also been used interchangeably.

4.1 Data collection and cleaning

4.1.1 Source selection and data extraction

Bibliometric analysis usually begins with a collection of publications that represents the main research focus of a field. The more complete and clean the dataset is, the less the noise, the more accurate the results, and the better the field is represented. Hence, to produce a comprehensive representation of DH, solid data collection and cleaning procedures are essential. In this study, the publication metadata were collected for both the ACA network and the co-authorship network.

The main data collection took place during the full month of September 2016 (from 2016-09-01 to 2016-09-30). As mentioned in chapter 3 (Methodology) all the peer-reviewed articles in the main issues of the three selected journals were collected. Following previous practices (e.g., Larivière et al., 2013; Nyhan and Duke-Williams, 2014a), content such as editorials, reviews, erratum, and notes were excluded because they were generally not peer-reviewed, nor considered as original contributions to the knowledge development. New articles published in these journals until the end of 2017 were added to the final dataset in January 2018.

The compiled dataset covers the whole publication periods of the three journals until 31st December 2017, i.e., *CHum* (1966–2004), *LLC/DSH* (1986–2017), *DHQ's* (2007–2017); none of these journals had publications that spanned the whole 52-year period (1966–2017).

When carrying out the collection, the essential items generally included the article titles, author names, author affiliations, publication year, journal name, author keywords, as well as the cited reference list. Each item was saved as a plain-text string/paragraph to fit into one cell of the metadata table (see Appendix A).

Although the task initially seemed trivial, it was surprisingly problematic and lengthy. At the time of writing (February 2019), 77.36% of the selected articles (1,955 out of 2,527) were either not indexed in Web of Science (or Scopus) or had incomplete

information.²² Besides this, indices have general problems that one needs to be aware of when downloading bibliometric data. Apart from limited coverage in Web of Science (MacRoberts and MacRoberts, 1996), this index is also criticised for its incomplete reference data (Zhao and Strotmann, 2015, pp. 72–74).²³ The coverage in Scopus, although wider than Web of Science, is relatively ‘short-term’ and ‘unstable’, and its indexing format is not consistent (Zhao and Strotmann, 2015, p. 78).

During the compilation stage, the task to fill out the missing data was more time-consuming than expected. Firstly, full-text PDFs were downloaded from the publisher and repository websites. Then, metadata was extracted from the PDFs as well as from the individual article webpages. Although a Python script was written and used for automatic webpage scraping, a significant amount of information on the PDFs needed manual checking.

CHum had many older articles as PDFs that were not OCRed or with low OCR quality (especially the range 1966-1990) and hence a straightforward copy-and-paste method could not get clean data without considerable manual typing work. Like *CHum*, references in some older *LLC/DSH* articles did not have online versions and also needed to be manually extracted. *DHQ*, on the contrary, is fully open access and born-digital from its launch, so the XML data was gathered from its website. However, the referencing styles in *DHQ* articles were not consistent, despite its author guidelines requiring a specific Harvard system. For example, references cited in (Crymble, 2016) followed APA style where all author surnames were put first and then followed by the year number, while references cited in (Svensson, 2010) followed Chicago style where only the first-author had surname put first and the year number was often at the end of the citation. The consistency in referencing styles is crucial in citation studies, especially when using an application to automatically identify the cited author and

²² For instance, *DHQ* was not included in Scopus, and despite the fact that the journal had sent articles to Web of Science since January 2017 (*DHQ*, 2017), only four volumes (2015-2018) were indexed so far. Similarly, the *LLC/DSH* articles published between 1986 and 2007 were not indexed in Web of Science, and its 2008-2018 volumes in Scopus did not have any reference data or full author names. The *CHum* articles in Web of Science lacked the volumes 1966-1967, and its 346 articles in Scopus had similar issues as *LLC/DSH*.

²³ For example, it only includes first-author names. Such data is not helpful when studying collaborative fields like DH where co-authored articles are common.

publication year. Therefore, heavy manual intervention was also carried out to convert all references to Parenthetical style, given that many articles were published as such.

Furthermore, all the references were kept in their original languages as shown in the published articles. It is assumed that the citation refers to the original work, which was written in that language, although no more than 200 references in the dataset were in languages other than English. Finally, the constructed dataset was saved as tables for further cleaning.

4.1.2 Data Cleaning and Formatting

It is essential in bibliometric studies that any problems with author names are dealt with first, otherwise they will cause obvious errors in the results. Allonyms (an individual having multiple names) and homonyms (multiple authors having the same name) are two common issues in bibliometric data, and they are often introduced by causes such as translations, transliterations, renaming, or errors in writing and indexing.

One study found 'nearly 9,000 unique economists, of whom 45% share a surname with at least one other economist in the data' (Goodman et al., 2015, p. 1393). Hence, cleaning the data, removing the duplicates, and formatting the styles are prerequisite procedures in author-based bibliometric studies, especially in DH where scholar names are understood to be more culturally diverse.²⁴

In the current dataset, the allonyms issue was the most frequent. Many authors had several name variations and some even more. For example, Willard McCarty had names such as 'McCarty, W', 'MacCarty, W', 'McCarthy, W', 'W Mccarty', 'Willard McCarty', 'Carty, W', etc., while Julia Flanders had duplicates like 'Flanders, j', 'fflanders, j', 'Flanders, J.H', 'J Flanders', and more. Generally, the longer and more complicated the name is, the more variations and duplicates it has. For instance, Lisa Lena Opas-Hänninen had name variations as 'Opas-hanninen, L.L', 'opas, I.I', 'Opas,

²⁴ Some notes need to be flagged when identifying authors. For example, in some cases, if the reference was to cite a book chapter, then the chapter's authors were credited instead of the editors of the whole volume. Conversely, if the whole book was cited, then the authors or the editors of the volume were counted as the authors for this citation. Additionally, where publications were translated, the authors, instead of the translators, were credited.

I', 'Opas-hänninen, I.I', 'Opas-Hanninen, I', and William A. Kretzschmar, Jr. had even more allonym issues, such as 'Kretzschmar, w.a', 'Kretzschmar jr., W.A', 'kretzschmar, w.a.jr', 'W.A Kretzschmar', 'William A. Kretzschma Jr', 'Kretzschmar, W', 'Kretzschmar, jr. W'.

The homonyms cases in the dataset were also problematic. For instance, 'Smith, J' had 62 citations in the dataset, which could have referred to John B. Smith at University of North Carolina, Joseph A. Smith at San Diego State University, or Joan M. Smith at the National Computing Centre, UK. Same surnames were very common among scholars, and some appear to be partners in life. For example, 'Siemens, R' and 'Siemens, L' were Ray and Lynne Siemens from the University of Victoria, Canada. 'Topkara, U' and 'Topkara, M' were Umut Topkara and Mercan Topkara at Purdue University, USA. All these names needed careful manual check against their original references and author affiliations.

To deal with such issues, some early studies proposed to distinguish the authors by retrieving their publication histories and creating a complete list of the bibliographies for each author (Eugene Garfield, 1979, pp. 243–244). This might be possible with a limited dataset of no more than 100 scholars but impractical to handle the 18,981 cited authors in this current dataset. Other studies suggested adding author affiliation information to assist the identification (Zhao and Logan, 2002), or to use an existing author ID system to distinguish scholars, such as Scopus ID (Tang et al., 2017, p. 990) or ORCID (Alonso et al., 2018, p. 3). However, using these systems also requires heavy manual work to identify and link IDs to names. Even if the names could be connected to their IDs, this approach was not applicable to the current dataset because most of the cited references were published long before the use of these ID systems and would not have them included.

To solve such problems, this study, firstly, converted author names to lower case to avoid duplicates with different cases. Secondly, because many authors only had their complete surname with first and/or middle names as initials, especially in the older content (e.g., *LLC* before 2000), all full names of the authors were manually checked to reduce any name disambiguation errors.

Following this, Strotmann's name disambiguation method was applied to help with the cleaning procedure (Strotmann et al., 2009; Zhao and Strotmann, 2011, p. 120), as well as Python scripts for repetitive batch editing and manual checking. Strotmann's method identifies two names as the same individual if the two names meet the following three requirements: a) the two names are mutually compatible; b) they share common co-authors; and c) they have not co-authored any papers together. A Python program was written according to b) and c); with the help of manual check on a), it was run six times until the program could not identify any more duplicates.

Finally, the formatted data of each journal was saved as an individual table and combined into one Excel file for further analysis (Appendix A). Each article record (or row) contained 42 data cells that included author name, affiliation, journal name, article title, publication date, abstract, funding information, references. The Excel file will be openly and freely accessible and downloadable from the institutional repository under a Creative Commons Licence after the completion of this PhD study. As it is problematic and time-consuming to construct a bibliometric dataset for DH, this dataset will be (at the time of writing) the first one available for reuse and will make a valuable contribution to other DH bibliometric studies.

From Appendix A, one can already see the general statistics about DH publications. Essentially, the number of articles collected each year fluctuates over time (see Figure 4.1).

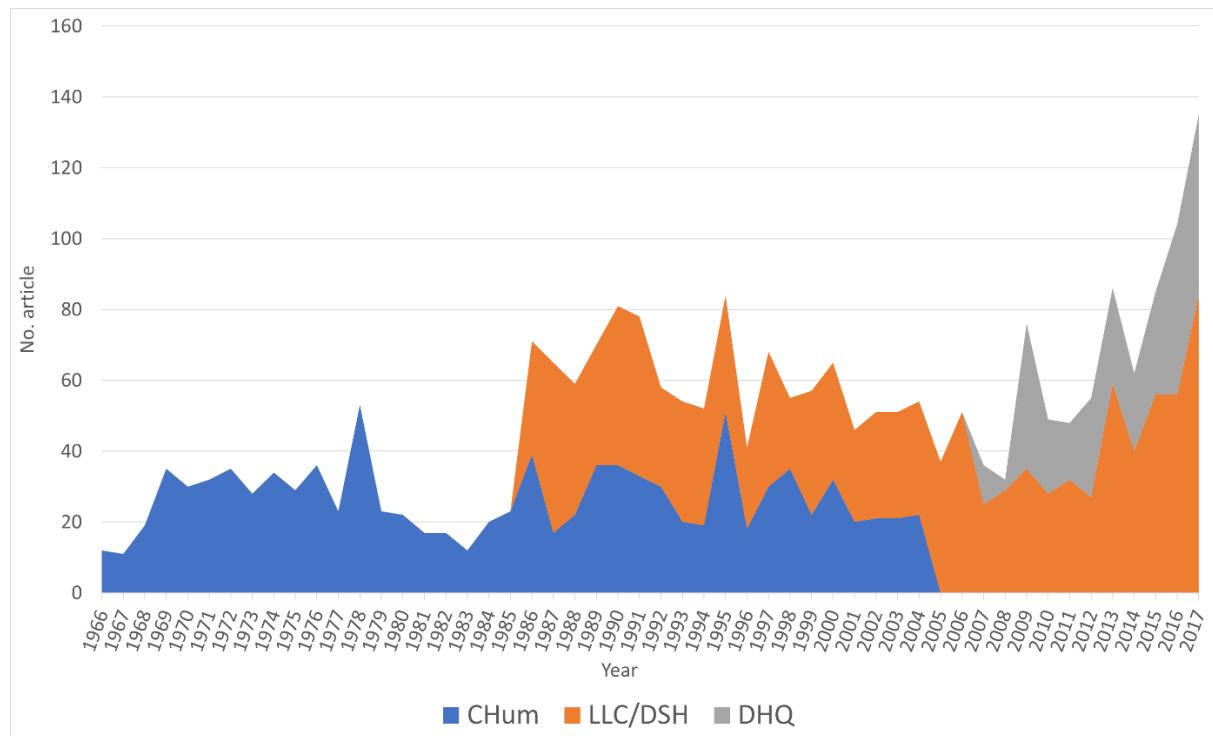


Figure 4.1: Area graph – the number of articles collected/published each year in journal *CHum*, *LLC/DSH*, and *DHQ* (1966-2017).

In total, 2,527 articles were collected (1,035 articles from *CHum*, 1,195 from *LLC/DSH*, and 297 from *DHQ*). The period between 1985 to 2017 had around 80.69% of the articles with two 10-year periods having the most articles, i.e., 1986 to 1995 (26.59%) and 2008 to 2017 (28.97%). The year 2017 had the highest number of publications, which was 135 (5.34%).

The overall number of citations has also been rising, especially in the past 10 years (2008-2017) where the number accounted for almost half (47.80%) of the total number of citations over the whole period (Figure 4.2).

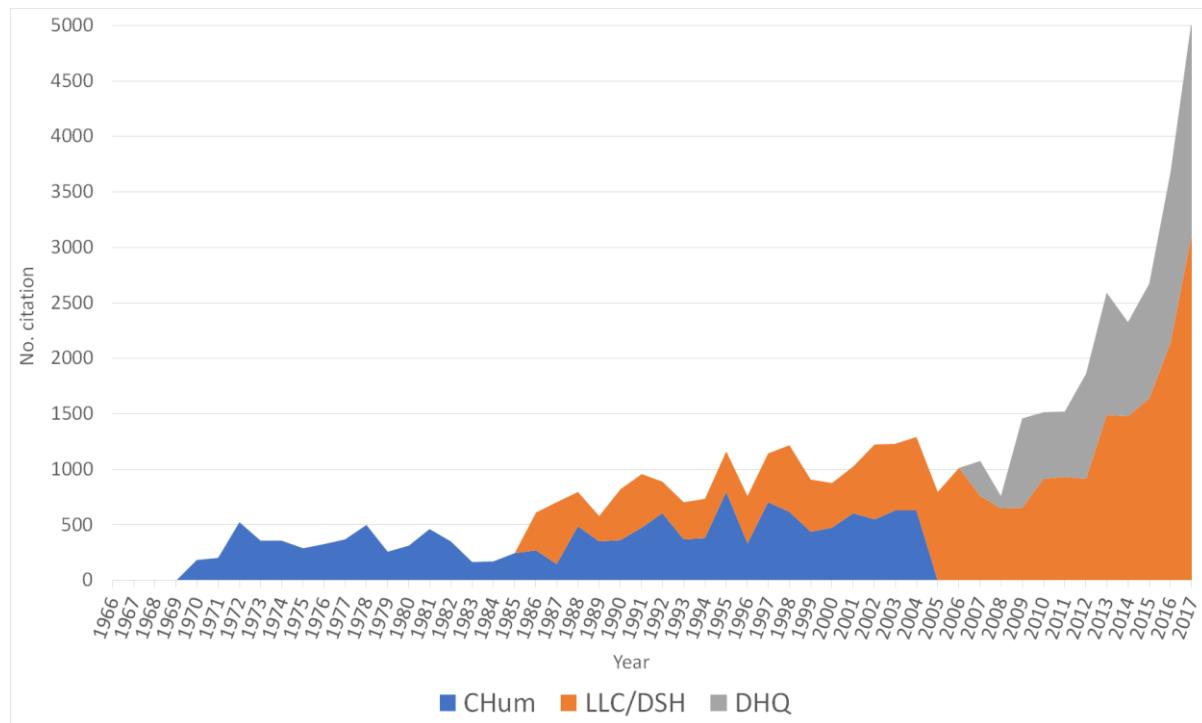


Figure 4.2: Area graph – the number of citations collected/published each year from *CHum*, *LLC/DSH*, and *DHQ* (1966-2017).

Figure 4.2 above shows that in total 49,047 cited references were collected (14,292 citations from *CHum*, 24,932 from *LLC/DSH*, and 9,823 from *DHQ*). As illustrated on the graph, scholars published in these journals tended to cite more and more references over time. This might be due to various reasons, such as, technological improvements, the ease of finding references, the advantage of open access and electronic publishing, and the usefulness of citation management tools. The graph below shows the average number of citations per article per year (Figure 4.3)

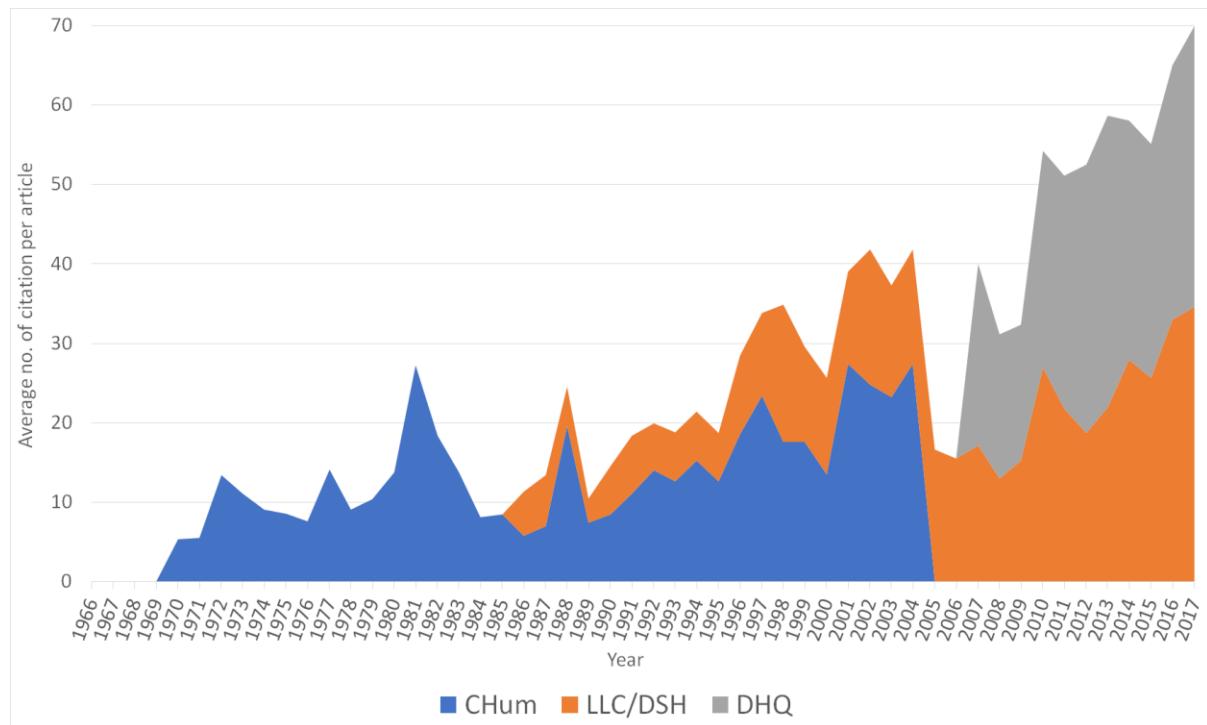


Figure 4.3: Area graph – the average number of citations per article collected/published each year for *CHum*, *LLC/DSH*, and *DHQ*.

The average number of citations per article varies across different disciplines. Some consider that hard sciences (e.g., medicine and biochemistry) often cited more references than the arts and humanities (e.g., literature, poetry and dance), as humanities used footnotes more frequently for explanations and so fewer citations are needed (Patience et al., 2017). Hyland, on the contrary, argued that humanities scholars take more space to express their research context with supportive references, while in the natural sciences, as the research context was well-known by its audience, fewer references were needed (Hyland, 1999, pp. 341–342). Another explanation may be that the more interdisciplinary the subject is, the more references it cites, because large block of texts and more references are needed to explain the relevant research backgrounds (Talja and Maula, 2003). Therefore, the growth in the number of citations in Figure 4.3 could potentially indicate the possibility of DH becoming more multidisciplinary, although further analysis is needed.

4.1.3 Further data collection

In order to understand the DH academic environment and discover its diversity as manifested in its publications, further data collection is needed. As discussed in

chapter 3 Methodology, this study specifically looks at two factors – gender and affiliated country (and language in Twitter networks in chapter 5), and how they influence the DH intellectual structure and scholarly communication. Related data can also be found in Appendix A, Appendix B, and Appendix E.

4.2 ACA network

The author co-citation analysis (ACA) network is the first of the four networks that this thesis constructs. In general, the ACA network involves two parts, the node and the edge. Authors were selected as nodes, and the co-citation relation between each two authors were calculated as edges (or links) in order to discover the connections between the authors. As mentioned in chapter 3 Methodology, using ‘fractional nonself citation count’ to select the top-cited scholars, and ‘exclusive co-citation count’ to calculate the co-citation matrix were the most preferred methods, especially for large-scale disciplinary studies (Ahlgren et al., 2003; Lindsey, 1980; Zhao and Strotmann, 2011). This study applied these two methods to construct the co-citation network and split the time into five periods to trace the longitudinal DH knowledge development. It then colour-coded the nodes according to their gender and country affiliation and calculated the betweenness centrality of the network.

4.2.1 Node

To conduct ‘fractional nonself citation count’, a unique list of all the cited authors in the current dataset was compiled. All authors were counted regardless of their by-line positions on the published document. In total, 18,981 unique cited authors were identified.

The total number of citations that each author received was calculated. For instance, if there were two articles that cited author A’s publications; article 1 cited three publications of author A, and article 2 cited four publications of author A, then the total number of citations that author A received was seven.

Following that, the number of self-citations was removed. Typically, a self-citation is defined as a citation in a publication of which at least one author (either the first author or co-author) is also the author of the cited publication (Noyons et al., 1999, p. 116); this type of self-citation is often called direct self-citation (Aksnes, 2003, p. 235).

Although self-citation is not something to be ‘condemned’ as it is usually fully appropriate to present fair and accurate previous work and it helps to avoid repeating what has already been studied elsewhere (Ioannidis, 2015, p. 10), this study followed the practice of the Web of Science to remove the number of direct self-citations from the total citation count of an individual author. This was achieved by checking if the author that was cited in an article was also the author (or co-author) of that article (if so, then the citations from this article were excluded when calculating the total citations this author received).

Finally, the fractional value was calculated based on the remaining numbers. As its name suggests, a fractional count calculates the citations of co-authored documents proportionally. This method was preferred for its emphasis on collaborative works (White and McCain, 1998, p. 327; Zhao and Strotmann, 2015, p. 28), and as it could ease the citation gap between the ‘newcomers’ and the ‘elders’ (Leydesdorff and Ophof, 2010, p. 2367). More clearly, when an article of n authors was counted, each

of these n authors’ fractional citation count equals to $\frac{1}{n}$.

By using ‘fractional nonself citation counting’, the values of the 18,981 cited authors were calculated and can be found in Appendix F. The following Table 4.1 shows the top 50 cited authors ranked by the fractional nonself citation value.

Table 4.1: The top 50 cited authors by the fractional nonself citation value, along with total citation and nonself citation counts, data from *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

	Author Name	Citation	Nonself Citation	Fractional Nonself Citation		Author Name	Citation	Nonself Citation	Fractional Nonself Citation
1	John Burrows	235	223	208.67	26	Allen H Renear	87	82	46.53
2	David Holmes	207	170	120.16	27	Geoffrey Leech	107	104	45.36
3	Jerome McGann	121	121	117.60	28	Andrew Morton	59	59	45.33
4	Peter Robinson	168	142	112.22	29	Matthew Jockers	60	55	44.57
5	Douglas Biber	132	130	99.07	30	Efstathios Stamatatos	64	61	42.12
6	Willard McCarty	131	105	89.06	31	David Yarowsky	59	55	41.25
7	David Hoover	118	90	88.17	32	Jay Bolter	59	59	41.03
8	Katherine Hayles	81	81	79.50	33	John Sinclair	51	48	40.51
9	MWA Smith	132	96	72.25	34	Patrick Juola	85	43	39.33
10	Johanna Drucker	82	80	69.13	35	Roberto Busa	45	45	39.25
11	Matthew Kirschenbaum	75	75	68.31	36	Noam Chomsky	41	41	38.33
12	Michael Sperberg-McQueen	134	128	67.28	37	Peter Shillingsburg	43	39	38.33
13	Lou Burnard	117	116	65.19	38	Friedrich Kittler	38	38	38.00
14	John Unsworth	78	78	61.41	39	Melissa Terras	81	62	36.37
15	Stephen Ramsay	64	64	59.95	40	Richard Forsyth	80	67	36.17
16	Susan Hockey	73	71	58.94	41	Julia Flanders	46	45	35.59
17	George Landow	65	65	55.33	42	George Yule	37	37	35.50
18	Franco Moretti	59	59	54.00	43	Frederick Mosteller	66	66	35.00
19	R. Harald Baayen	101	93	51.48	44	Donald mckenzie	51	47	34.50
20	Hugh Craig	69	64	51.33	45	Shlomo Argamon	84	74	34.20
21	Lev Manovich	50	50	50.00	46	Stig Johansson	62	62	34.13
22	Alan Liu	53	53	48.56	47	Tom Merriam	75	43	34.00
23	Espen Aarseth	47	47	47.00	48	Maciej Eder	66	43	33.83
24	Michael Halliday	55	55	46.90	49	Geoffrey Rockwell	61	49	33.58
25	Gregory Crane	101	69	46.79	50	Charles Martindale	65	47	33.28

In Table 4.1, the values of the three counting methods vary significantly. For example, Katherine Hayles was the 8th most cited author with 81 citations, and as she never published any articles in the collected journals, she did not have any self-citations to be removed. Most of her articles were single-authored, therefore, her fractional nonself citation value was 79.5. Conversely, Michael Sperberg-McQueen had 134 citations,

but because most of his articles were co-authored, he was ranked 12th with fractional nonself citation value of 67.28.

4.2.2 Edge

The edges of co-citation define the relationships between nodes and form an important part of the author co-citation network. ‘Exclusive co-citation count’ is counted when an article cites at least one paper of author A and at least one of author B that is different from the one of A (in any author position); here the ‘exclusive co-citation count’ of authors A and B increases by 1.

Each cited author was in turn paired with the rest of the 18,980 authors to calculate the occurrences that each pair was cited together. Given the nature of co-citation counting, the edges were undirected. Table 4.2 shows an example of the author co-citation matrix of the top eight ranked authors.

Table 4.2: Author co-citation matrix of the top 8 cited authors, data from *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017

	John Burrows	David Holmes	Jerome McGann	Peter Robinson	Douglas Biber	Willard McCarty	Katherine Hayles	David Hoover
John Burrows	0	205	4	3	17	18	2	169
David Holmes	205	0	0	2	31	9	0	81
Jerome McGann	4	0	0	29	1	62	25	8
Peter Robinson	3	2	29	0	1	8	2	1
Douglas Biber	17	31	1	1	0	0	0	14
Willard McCarty	18	9	62	8	0	0	12	12
Katherine Hayles	2	0	25	2	0	12	0	1
David Hoover	169	81	8	1	14	12	1	0

As can be seen in Table 4.1 above, author John Burrows and David Holmes were co-cited by 205 articles in the citation dataset, while author John Burrows and author Jerome McGann were only co-cited by 4 articles. This means, Burrows had a much closer research interest with Holmes than with McGann, and should be placed closer to Holmes than to McGann on the visualised network. Because this study applied

exclusive count, the diagonal values on the table were all zero (e.g., where co-citation value between author A and themself was excluded).

4.2.3 Network Visualisation

In order to visualise the co-citation network more efficiently, this study has selected 3,477 most cited authors out of the 18,981 authors as network nodes (those whose nonself citation value is greater than, or equal to, 2). This selection was made for efficiently reducing the visualisation software running time, which was necessary given the scope of this PhD study and the various networks to be constructed. By reducing the nodes to 3,477 on the network, it also helps to compare it with other networks in this thesis at similar scales. Although limited, 3,477 is already larger than many data samples of other ACA studies published in recent years. For example, only 100 most cited authors were selected in Bu's, Jeong's and Hsiao's ACA network visualisations (Bu et al., 2016; Jeong and Song, 2016; Hsiao and Chen, 2017) while Kim selected 669 and 967 authors for their two network visualisations (Kim et al., 2016).

The citation matrix was imported into the network visualisation software, VOSviewer 1.6.8, a free software for constructing and viewing networks. There is a recent trend towards ACA visualisation studies to construct larger maps (van Eck et al., 2010, p. 2406), yet, simple graphical representations constructed by software like SPSS or Pajek are becoming inadequate to yield satisfactory results when the dataset is larger than 100 nodes. Gephi is also a popular choice and was used in this study for centrality measure (explained in section 4.2.4 and 4.3.4), but the labels on its network become messy and sometimes, unreadable when they overlap. VOSviewer is good at handling large-scale datasets, and it employs algorithms to optimise the label display and overlapping issue (van Eck and Waltman, 2010, p. 530). This software is not only good for bibliometric visualisation, but also employs many mapping techniques such as multidimensional scaling (Borg and Groenen, 2005, pp. 1–3), and VOS (van Eck and Waltman, 2007, p. 299), and VxOrd (Klavans and Boyack, 2006, p. 251) that can construct other types of social network visualisations. It has many functions to enhance the graphical representation of large maps, such as zoom functionality, labelling algorithms, and density metaphors (van Eck and Waltman, 2010, p. 524).

More specifically, the network visualisation procedure using the VOSviewer consisted of four steps. The first was to convert and normalise the author co-citation matrix into a similarity matrix with association strength as the similarity measure (or proximity index) (Peters and van Raan, 1993, p. 5). This step was known to be more useful than the other measures like cosine or the Jaccard index (Eck and Waltman, 2009, p. 1637) for bibliometric analysis.

Secondly, a two-dimensional map was constructed from the similarity matrix, and the distance on the map between any pair of authors A and B was their similarity s_{AB} . Authors that had higher similarities were located more closely, while lower similarities were further apart. Problems occurred when calculating the minimised sum of the squared Euclidean distances between pairs. Given that the higher the similarity values, the greater the weight of their squared distance, and most of the authors would have 0 as the coordinates to achieve the minimised value, their locations would all be the same, e.g., (0,0), on the constructed map. In order to avoid such a useless map, VOSviewer employs a method to keep the average distance between two authors to 1. The equations are as the following ($n = 3477$)

$$V(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{A < B} s_{AB} \|\mathbf{x}_A - \mathbf{x}_B\|^2 \quad (2)$$

where author A 's location value \mathbf{x}_A was a vector which contained the coordinates $\mathbf{x}_A = (x_{A1}, x_{A2})$, and $\|\mathbf{x}_A - \mathbf{x}_B\|$ was the Euclidean norm. The goal was to find out a set of \mathbf{x}_A that could calculate the minimised value of the equation.

$$\frac{2}{n(n-1)} \sum_{A < B} \|\mathbf{x}_A - \mathbf{x}_B\| = 1 \quad (3)$$

where the average distance of the 3,477 selected authors was 1.

It should be noted that the position and distance were not linear. Take the earlier example, author Burrows and Holmes were co-cited by 205 articles in the citation dataset, while author Burrows and author McGann were only co-cited by 4 articles. The ratio of the distances between author Burrows to Holmes and Burrows to McGann was not 4:205 (i.e., not linear).

Thirdly, the citation network was then translated, reflected and rotated. To optimise the network's centre to the correct location, each coordinate was translated, and reflected in the vertical axis if the median of x_{A1} was greater than 0, or in the horizontal axis if the median of x_{A2} was greater than 0 (van Eck and Waltman, 2010, p. 532). The rotation employed principal component analysis (PCA) to maximise the variance on the horizontal dimension, so that the network could be positioned at the same optimised location each time of the visualisation.

Finally, the author nodes on the network were clustered by a unified approach (Waltman et al., 2010, p. 630) as the following ($n = 3477$)

$$d_{AB} = \|\mathbf{x}_A - \mathbf{x}_B\| \quad (4)$$

$$d_{AB} = \begin{cases} 0 & \text{if } \mathbf{x}_A = \mathbf{x}_B \\ \frac{1}{\gamma} & \text{if } \mathbf{x}_A \neq \mathbf{x}_B \end{cases} \quad (5)$$

d_{AB} represented the distance between authors A and B , and the parameter γ was the resolution parameter ($\gamma > 0$) which was given. The default value of γ was 1 in the VOSviewer. The greater the value of this parameter, the more the number of clusters.

$$V(x_1, \dots, x_n) = \sum_{A < B} s_{AB} d_{AB}^{-\gamma} - \sum_{A < B} d_{AB}^{-\gamma} \quad (6)$$

This equation was used to calculate the attractive and repulsive forces between authors by finding a set of x_n so that (6) was minimised, then each x represented a number of the clusters. The first part on the right was to calculate the attractive force, and the second was the repulsive force. Thus, the final force between two authors was the result of the attractive force subtracted by the repulsive force. According to the equation, the higher the association strength s_{AB} , the stronger the attractive force between the authors, but the repulsive force did not relate to the s_{AB} value, so the general result of the two forces was that authors with higher s_{AB} were pulled closer while with lower s_{AB} were pushed further apart. By applying this clustering method, even small clusters could always be identified by selecting a large import of the parameter γ .

The network visualisation is shown in Figure 4.4 below. The size of the node (or circle) represents the nonself fractional citations this author received according to the dataset, and the higher the fractional citation value, the larger the node is. It shows a clearer picture of the DH intellectual structure, and results will be further discussed at the end of this section (4.2.6 Discussion and analysis).

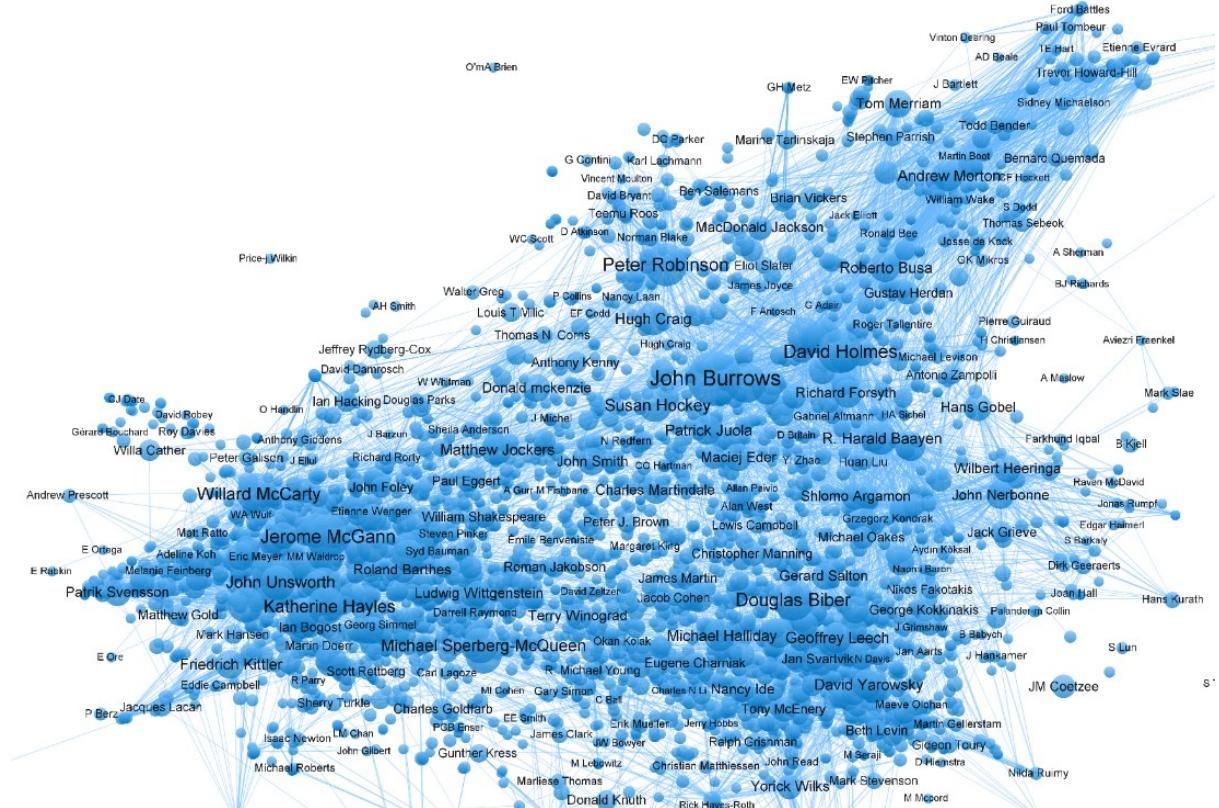


Figure 4.4: Author co-citation network in DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

It should be noted that while a network as a means can present direct relationships of research subjects, the interactive graphs that VOSviewer generated can only be viewed by installing the software. On the screenshot (e.g., Figure 4.4), many author nodes were blocked by larger nodes and could not show the details. To address such issue, this study has used SigmaJS²⁵ to publish the networks online, and this not only gives more freedom to build the network visualisations, but also allows other scholars to explore the details of the networks more easily (e.g., by filter and search functions).²⁶

4.2.4 Centrality

When analysing and interpreting the visualised network, centrality analysis is often applied to gain more insights from the visualisation (Marsden, 2002). It consists of

²⁵ SigmaJS, a JavaScript library, for more information <http://sigmajs.org/>.

²⁶ The online version can be found from: http://jin-gao.com/map/view.html?citations_all

many practical methods that are often called the centrality indices, such as *betweenness centrality*, *degree centrality*, or *closeness centrality* (Freeman, 1978). These methods quantify the shortest paths between pairs of nodes to interpret the structural prominence and the social importance of the individuals based on their node positions in the network (Koschützki et al., 2005).

The general definition of a certain node's centrality is quantified by the range of this node to the others that are connected to it within a given network (Wasserman and Faust, 1994, pp. 178–218). Among the three main centrality indices, *betweenness centrality* usually explores the control and mediation of the node; *degree centrality* often analyses the direct links of the node; while *closeness centrality* commonly researches the flow of information of the node (Badar et al., 2013, p. 758). Some claimed that the higher the centrality, the greater the research impact of the scholars was in the network (Yan and Ding, 2009, p. 30).

This study has employed the *betweenness centrality* approach for all four networks as this approach is designed for indicating bridging nodes (Lu and Feng, 2009). It measures the average number of times each node acts as a bridge along the shortest possible path between any pair of other nodes (Brandes, 2001). In other words, it counts the number of the shortest possible paths of any other two nodes passing (or 'between') a node (Badar et al., 2013, p. 759), i.e., the higher the *betweenness centrality* value of a node, the more pairs that have the shortest paths to go through this node, and the more central and important its position.

This method was applied to study the degree of control and influence a person could have upon the network of communication with other people as early as 1977 (Freeman, 1977). Studies have shown that nodes that have higher betweenness centrality might be more likely to have 'instrumental outcomes', such as productivity and creativity (Perry-Smith and Shalley, 2003; McFadyen and Cannella, 2004).

Gephi 0.9.2, an open-source network analysis and visualization software package, was used to calculate the *betweenness centrality* (Brandes, 2001), and results have been saved in Appendix F and will be discussed later at the end of this section (4.2.6 Discussion and analysis).

4.2.5 Five Periods

It was effective to explore the general intellectual structure and community from the general co-citation network. However, as DH has been changing and developing overtime, a single synchronic ACA representation would not be enough to mirror the everchanging disciplinary development. By splitting the period into different stages, it is more effective to find how the field was formed, and how its history has developed. The longitudinal approach not only traces the trajectories of DH development over time, but also explores the steps that individuals took to reach their current position in the network (Tang et al., 2017, p. 991).

To examine the evolution of DH, this study has divided the 52-year bibliometric data period into five periods. For the rationale of division, this paper followed the historical stages that Susan Hockey discussed in 2004 (Hockey, 2004) where she split the history of humanities computing into four periods: 1949 to early 1970s (Beginnings), 1970s to mid-1980s (Consolidation), mid-1980s to early 1990s (New Developments), and early 1990s to the present (the Era of the Internet). It was a very clear segregation supported by many ‘highlighting landmarks’, but up to the time of writing, the last period ‘the Era of the Internet’ had been around 29 years, which counts for more than half of the citation data period. Taking this into consideration, this study cut the last period in half by the year of 2005, being the time in which the term Digital Humanities had become widely established as proposed by (Nyhan and Flinn, 2016, pp. 1–3).

Therefore, according to the citing articles’ publication dates, this study divided the 52-year citation data into five stages: 1966 – 1970, 1971 – 1985, 1986 – 1990, 1991 – 2005, 2006 – 2017. As mentioned in the earlier sections, around 3,000 of the most cited authors in each time period were selected for the network visualisation, although some periods do not have that many nodes.

Visualised networks are shown as the following and further analysis on the network are at the end of this section (4.2.6 Discussion and analysis).

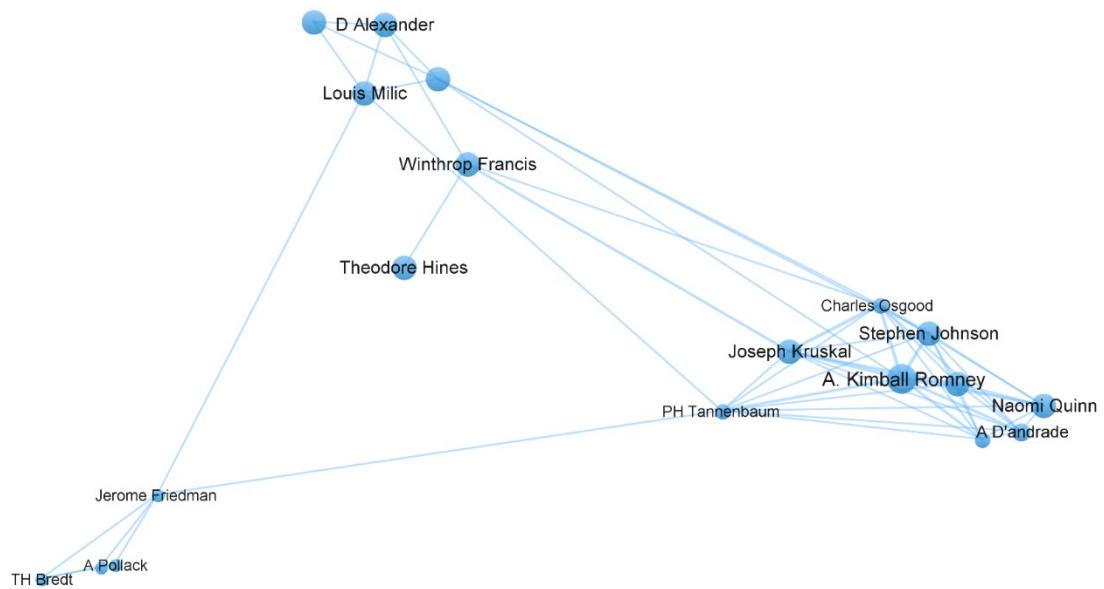


Figure 4.5: Author co-citation network of DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-1970.

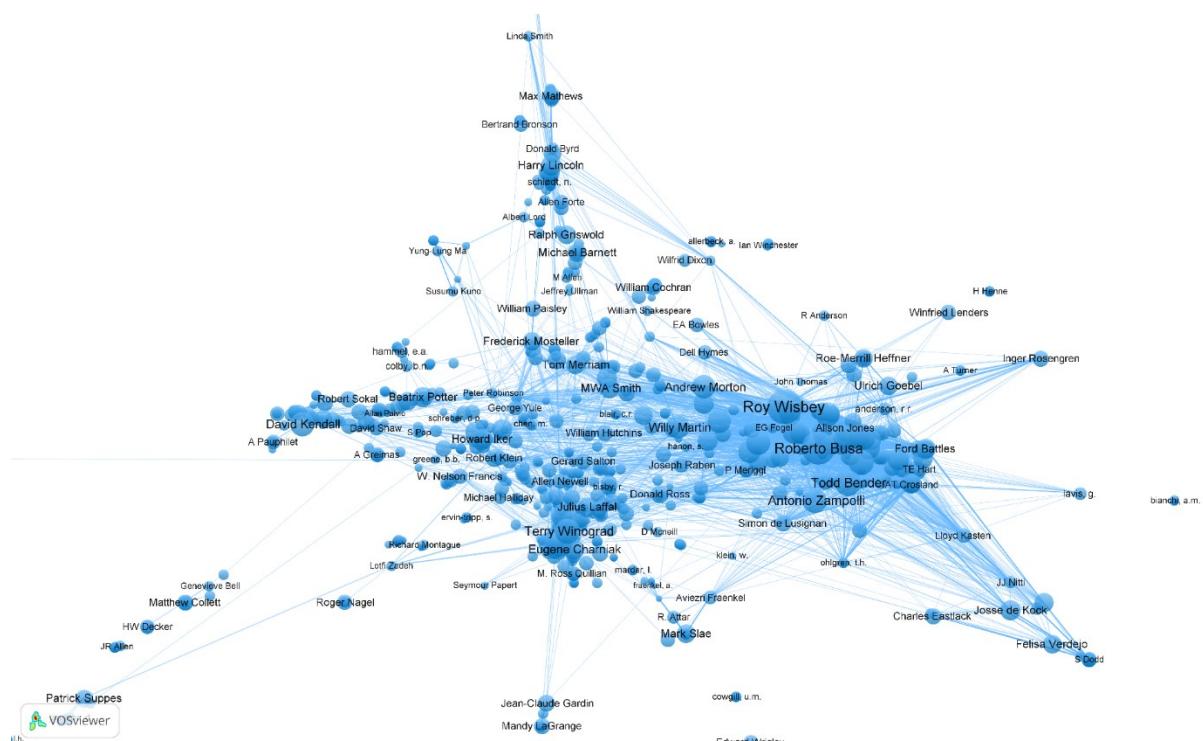


Figure 4.6: Author co-citation network of DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1971-1985.

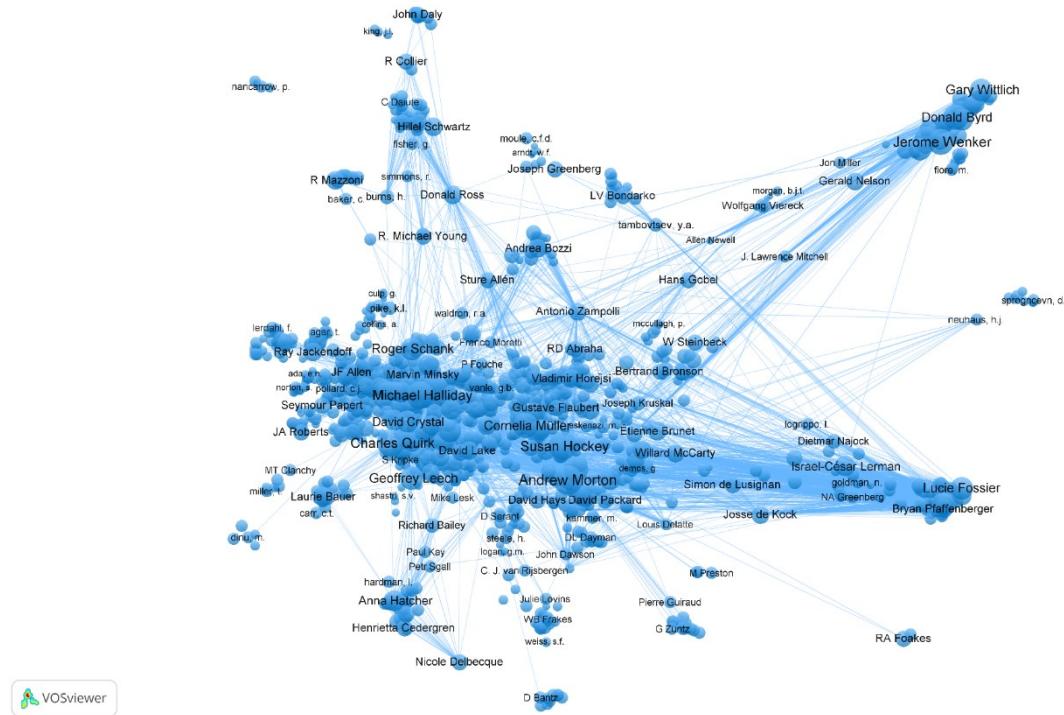


Figure 4.7: Author co-citation network of DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1986-1990.

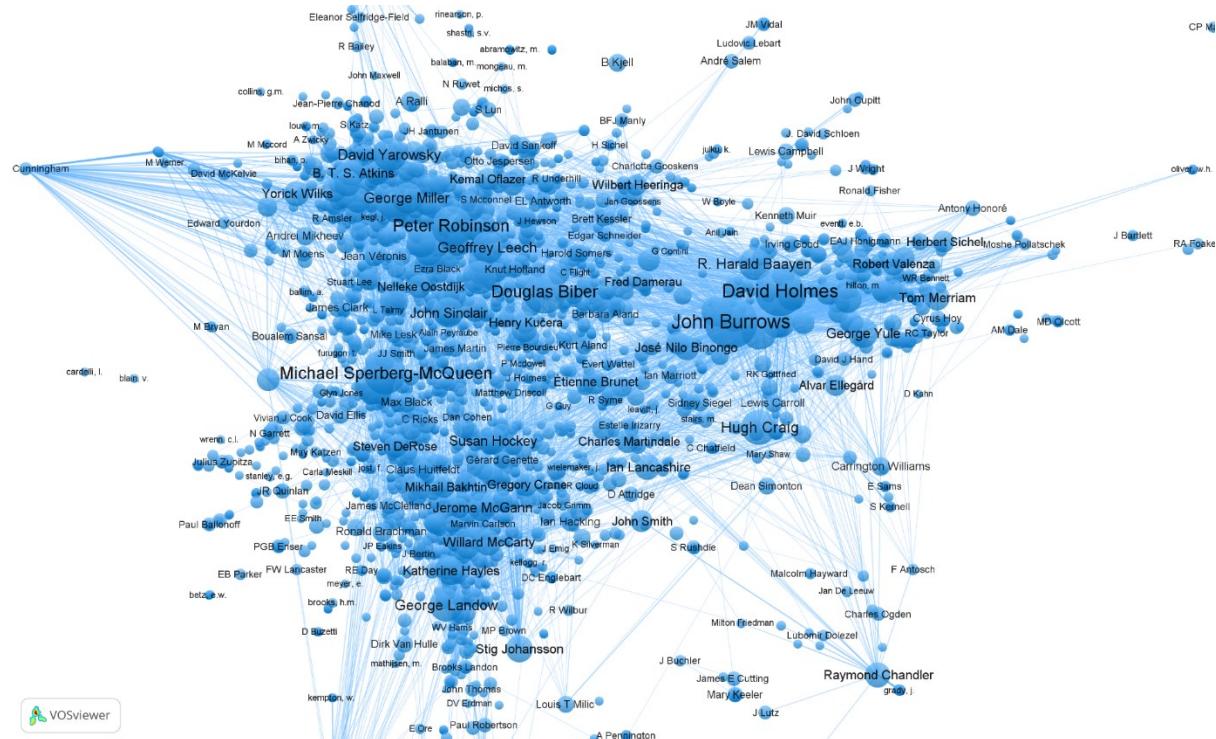


Figure 4.8: Author co-citation network of DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1991-2005.

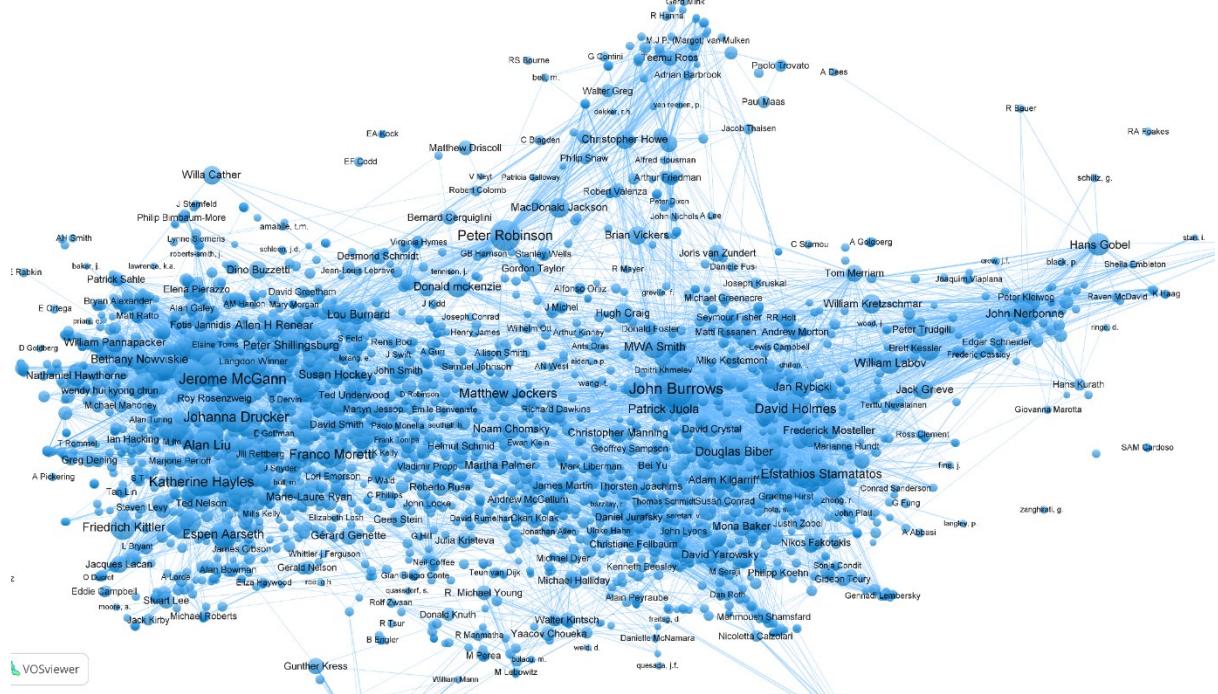


Figure 4.9: Author co-citation network of DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 2006-2017.

4.2.6 Discussion and analysis

This section presents a series of discussions to interpret and analyse the ACA network. Such interpretation will also appear at the end of the sections of the other three networks.

Interpretation, in this study, is referred to as the construction of meaning and understanding through previously visualised networks, aiming to produce 'an understanding of the context of information systems, and the process whereby the information system influences and is influenced by its context' (Howcroft and Trauth, 2005, p. 83). Interpreting networks not only helps scholars to revisit research topics and questions from a broader view, but also assists in redefining and reviewing the subjects that networks represent (Krieger and Belliger, 2014, p. 8).

Although network visualisation is powerful in displaying complex data, it should be noted that different methods perceive network structural features differently (McGrath et al., 1997, p. 223). Interpreting nodes and edges without context may misrepresent the structural properties constructed through quantitative approaches. This study, therefore, cautiously deciphers what the networks reveal with evidential support from

previous scholarship. It starts this process by summarising the results, looking for general explanations of the networks, and discussing insights reflected on the general DH communities.

It should be noted that although this study does not focus on identifying individuals, in order to give examples and provide interpretations of network results, a few scholars and their works are named and discussed. This demonstration approach is common in network studies, and has been applied by many scholars, e.g., (Wang and Inaba, 2009a; Lee et al., 2017). As mentioned in the Methodology chapter (section 3.4), this study has been approved by the departmental research ethics adviser, and it follows UCL's data protection regulation and uses only published information on academic literature.

4.2.6.1 The ACA network structure

Robertson described the DH subject structure as a house with many rooms:

Moving forward, we would be better served by reimagining digital humanities not as single all-encompassing tent but as a house with many rooms, different spaces for disciplines that are not silos but entry points and conduits to central spaces where those from different disciplines working with particular tools and media can gather. Each of the many disciplinary rooms would have a distinctive character, reflecting a particular contribution and orientation to the field. (Robertson, 2016, pp. 290–291)

By reviewing the results from the co-citation network, we can see that the DH intellectual structure partly agrees with Robertson's 'house structure' although there are not many rooms discovered. Yet, there is one particular 'room' that this study labels as 'general historical literacy and information science', and it does not fit into any specific discipline but rather represents a mixture of several intertwining fields.

On VOSviewer, there are two ways to show the network, one is a normal network visualisation (Figure 4.4 in section 4.2.3), the other is a density (or heat map) view where clusters are shown more clearly (Figure 4.10 below).²⁷

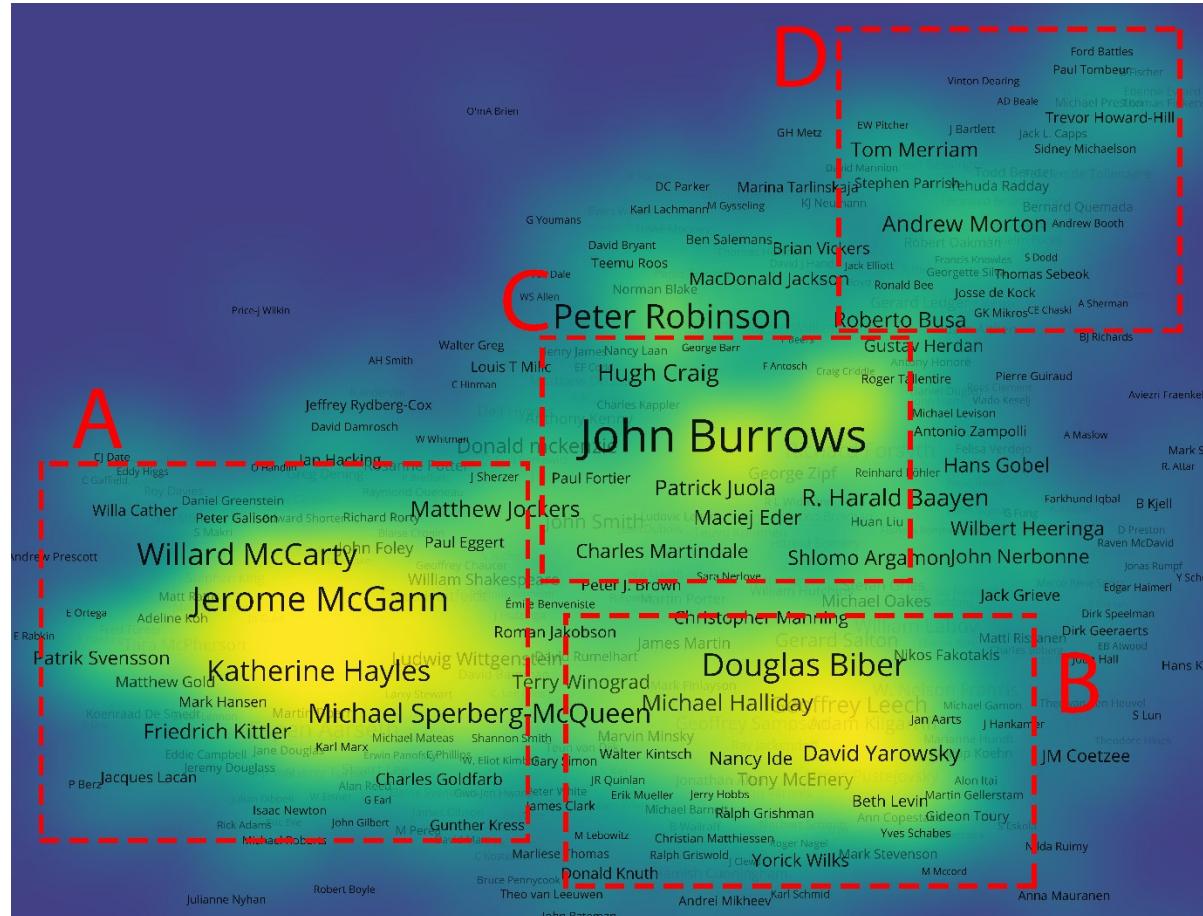


Figure 4.10: Density view of author co-citation network in DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

As shown in Figure 4.10, the DH ACA network consists of four loosely connected clusters²⁸ – general historical literacy and information science (cluster A, on the left);

²⁷ On the density map, the higher the density value of the node, the more yellow than blue the colour was. The density value of the node depended not on the node itself, but on the size, number and distance of the nodes around it. That is to say, the larger the size, the greater the number, and the closer the distance of the nodes around this node, the higher the density value was.

²⁸ In this thesis, the cluster boxes drawn on all the networks are only for the convenience of demonstrating estimate positions on the network, and there is no direct indication of inclusion or exclusion of particular nodes or names. In addition, because there is no standard category taxonomy to help name the network clusters presented in this study, cluster names were given as the most understandable format, and there are also general explanations in later sections to assist the understanding and discussion of the clusters.

computational linguistics (cluster B, on the bottom right); English studies (cluster C, in the middle); and the studies of early projects and pioneers (cluster D, at the top right). The clusters are not labelled as one type of category. As there is no standard category taxonomy, this study has named the clusters as the most understandable format to better assist the interpretation and discussion.

3,477 scholars are positioned in one (or more) of those four clusters according to the co-citation relationship based on the collected journal data. Although, through our understanding, many scholars do not belong exclusively to a particular cluster (some might even argue that they do not belong to any of the clusters), this study draws general conclusions that are mainly reflected in the visualised networks from a macro perspective. As DH is highly interdisciplinary, it is challenging and sometimes impossible to assign DH scholars into a specific category. It is noted that errors and disputes can be introduced during the discussion; they can be unpacked in further studies, and ideally in a case-by-case arrangement to study them individually and more thoroughly.

As the network result shows, DH in general has developed into a mature status that consolidates different origins to one integrated territory, i.e., the process of integration (Porter and Rafols, 2009; Rafols and Meyer, 2010). Subjects that are believed to belong to a few disciplines can be seen on the network, such as Humanities (e.g., English studies, historical literature), Computer Sciences (e.g., natural language processing), and Applied mathematics (e.g., statistics). During the process of integration with other disciplines, DH has absorbed and exchanged research themes, approaches, as well as outputs (Wagner et al., 2011). It seems that cluster A is the core of DH, or at least, the contemporary core, while others are subgroups related to DH that have been moving away from the core (based on the longitudinal view, see section 4.2.5 Five Periods).

4.2.6.2 Cluster A – General historical literacy and information science

The largest cluster that contains the most cited DH authors (42.4%) is related to topics of general historical literacy and information science, and it is not one but multiple intertwining fields. Some of them share field names or overlap with each other, such as digital history, digital literacy, digital media studies, historical studies, information

literacy, historical literacy, information studies, quantitative history, library science. In general, these fields can be grouped into three main areas – history studies, information studies, and new media studies, and additionally, a very small number of overlapping literature studies.

Firstly, history studies in DH (or sometimes called ‘digital history’ although distinctions exist) is one of the popular topics and it continues to attract interest from a variety of backgrounds (Brennan, 2018). It applies digital methods and tools to facilitate historical analysis, presentation, and research. Some believe that it is a branch of DH, while others think the two have overlapping as well as excluding areas (Zaagsma, 2013).

Secondly, information studies (or information science) is a field that is generally interested in information, such as information retrieval, collection, classification, manipulation, storage, dissemination, and protection (Stock and Stock, 2013, p. 3). In the past, information science used to be associated with fields like computer science, telecommunications, and psychology (Yan, 2011, p. 510). However, according to Yan, lately, information science has been absorbing interests from a variety of areas, and many of them are related to arts and humanities, e.g., archival studies, linguistics, and museology. Many DH studies support the idea of information studies being an obvious component in DH. For instance, as mentioned in section 2.1 (DH subject specialty), Salah and other authors explained the apparent presentation of library and information science in their DH network study (Salah et al., 2015, p. 83). Robinson et al. and Koltay also argued that there was partial integration between the two, and they believed this integration would bring a positive future for both sides (Robinson et al., 2015; Koltay, 2016).

What is the relationship between history studies and information studies in DH? Robertson argued that there are distinctions between the two, as they have different tools, methods, and values (Robertson, 2016). However, they share common research interests that are closely related to each other and to DH. There is an area where these two fields meet – ‘historical information science’ (Boonstra et al., 2006), and it attracts the majority of DH scholars forming a very dense cluster on the network. Boonstra pointed out that ‘historical information science’ was once called ‘e-

humanities' and 'humanities computing' (Boonstra et al., 2006, p. 13), which reflects an apparent relationship to digital humanities.

Based on the current dataset, many cited studies in cluster A are from this intersection. For example, the book by Jerome McGann (a textual scholar) is cited 37 times which is the most cited publication in cluster A ([McGann, 2001](#)). 'Humanities Computing' by Willard McCarty (a Professor of Humanities Computing) is cited by 35 articles in the dataset (McCarty, 2003a), and is one of the earliest theoretical works on modelling and is widely regarded as one of the foundational commentaries on Humanities Computing and Digital Humanities. The book by Franco Moretti (a literary scholar) is cited 25 times (Moretti, 2005), and as a 'great iconoclast of literary criticism', it provides abstract models for studying literary history (Sutherland, 2006). Katherine Hayles' book (Hayles, 1999) has 21 citations in the dataset, ranging widely across the history of technology, cultural studies and literary criticism. It is difficult to tell if these publications are devoted to history studies or information studies, as they are interdisciplinary and make contributions to both.

Thirdly, there are many scholars doing new media studies in cluster A. For example, the book by Matthew G Kirschenbaum (Kirschenbaum, 2008) has been cited by 29 articles in the current dataset, Lev Manovich (Manovich, 2001) has 28 citations, and the book by Jay David Bolter and Richard Grusin (Bolter and Grusin, 2003) is cited 26 times. These works can also be categorised into history studies and information studies as they not only review the historical development of new media but discuss it as a field of a new form of information.

New media (or sometimes called 'digital media') are forms of media that differentiate themselves from 'old media' as they are native to computers and relying on computers for redistribution (Wardrip-Fruin and Montfort, 2003, p. 13). Scholars started to debate the relationship between new media and DH from 2010 (Fitzpatrick, 2010). As Fitzpatrick discussed, new media seemed to be a new focus of DH only in recent years, and it was not something that the field had been doing in the past. She also made the point that 'institutional turf wars' and distinct institutional structures caused these debates about the relationship between new media and DH:

Some of this debate [about the definition of the Digital Humanities and its relationship to digital media studies] arose, I think, from a sense of annoyance among folks who've been working in DH for years that suddenly, now, with the rise of social media and the visibility of those working in and on those forms, a bunch of attention is being paid to something called 'digital humanities' — but the thing going by that name isn't quite the same thing that it's been for the past few decades, and the thing that DH has been is now being overlooked (or worse, dismissed) in favor of this new interest in digital media. (Fitzpatrick, 2010)

To many, DH is an academic discipline while new media is an industry or category of technology that can be analysed by many types of studies, but it is also believed that there is a substantial overlap between the two areas. For example, after comparing 50 syllabi from the two areas, Stutsman found that they shared the majority of modules, signifying a strong relationship and high similarity (Stutsman, 2013).

Besides this, a small number of the works mentioned above are also related to literature studies, and most of them are interdisciplinary works ranging across various fields, such as (Moretti, 2005) and (Hayles, 1999). Although literature studies are sometimes believed to have a close relationship with English studies (especially in an English-speaking context) (McMurtry, 1985), these works mentioned are written from a more general perspective to discuss literature studies as a field, instead of analysing specific English literature, and thus, they are different from the works which appeared in cluster C – English studies. In addition, we could also find many other disciplines mingled with the three main areas within cluster A, such as Classics (e.g., Coffee et al., 2013), archaeology (e.g., Eiteljorg, 2004; Forte, 2015), music (e.g., Fujinaga and Weiss, 2004; Burgoyne et al., 2015), and image processing (Terras, 2012b).

Why does cluster A account for the largest portion in DH? Salah et al., tried to explain it as the 'Digital Humanities' term usage in publications of the humanities disciplines (Salah et al., 2015). They argued that when we study DH, most bibliometric datasets were compiled by searching the names of the field (i.e., 'Digital Humanities'), and they were thus more related to the titles found in the history, new media, and library and information studies. In these fields, 'Digital Humanities' are often treated as a research subject to study its history, the new media involved, and its disciplinary status. Because these DH disciplinary topics are common in these three fields (such as

knowledge structure, intellectual boundaries, and disciplinary history), the term ‘Digital Humanities’ is used very often in their publication titles when compared to other disciplines which can result in a high frequency of occurrences when conducting bibliometric analysis. This potentially explains why there were more representations in DH from these fields and scholarly activities in DH were relatively limited to these fields according to the current dataset.

Nevertheless, the close relationship between DH and the fields in cluster A is not only found in bibliometric data, but in many other studies (Robinson et al., 2015). There are similarities between DH and subjects in cluster A in many ways that are beyond the limitation of mere bibliometric dataset construction. For example, they share a common focus on the study and practice of recorded information (e.g., documents, archives), and thus, Koltay considered them all to be ‘sciences of information’ (Koltay, 2016, p. 782). All of them are interdisciplinary fields addressing topics across humanities, social sciences, and applied sciences, and all of them have emerged partially from service functions associated with the academic use of recorded information (e.g., library, archive, museum). They all have links between their academic disciplinary status and a support role for research (Robinson et al., 2015). Such links also introduce identification and definition problems in all subjects (Yan, 2011, p. 510; Warwick, 2012, p. 193). In addition, many DH centres are housed within the department of information studies (e.g., UCL in the UK²⁹, Wuhan University in China), the department of history (e.g., Humboldt University of Berlin, Cleveland State University in the US, and Nanjing University in China), or department of media studies (e.g., MIT Comparative Media Studies, Beijing Mobile Media and Cultural Computing), although other disciplines are also involved, e.g., English department (see section 4.2.6.4 Cluster C – English studies). Educational materials as well as journals of all subjects share many similar contents, tools, and methods (Robinson et al., 2015).

²⁹ UCL Centre for Digital Humanities (UCLDH) was founded in 2010. At that time, positioning a DH centre within the Department of Information Studies was seen as unusual. According to the candidate’s supervisor, it used to be said that it was one of the few DH centres in the world, apart from University of Illinois Urbana-Champaign that was founded within a Department of Information Studies. However, ten years later, we see this has become more and more common in many other universities.

Such institutional and pedagogic settings indicate a clear shared focus among organisations from both DH and the subjects in cluster A.

On the other hand, these areas also have differences, although many have pointed out that there is substantial overlap, and scholars in historical information science even traced their ‘founding father’ to Roberto Busa, too (Boonstra et al., 2006, p. 25). An apparent distinction is that history studies is genuinely rooted in the Humanities, while information science is believed to be one of the social sciences (Hjørland, 2000), and new media is more related to technology, despite the fact that many of their research questions and approaches overlap with the Humanities (Head, 2008).

It seems that the subjects in cluster A have been developed and integrated into an area where DH values (e.g., interdisciplinarity, collaboration) are being celebrated. Just as Kathleen Fitzpatrick once described her experience:

I've operated on the edges of a number of disciplines, fully fitting into none of them, but trying to find ways to get them to talk to one another. So I'm sort of a literature person, but stay on the fringes of the discipline. I've had good conversations with folks in American studies, though my work doesn't have the interest in the historical, the national, or the hemispheric that would make me genuinely a member of that field. I've been to half a dozen Internet Research conferences, and have loved them, but my methodologies are vastly different from those of most of that group's members. And I'm clearly in media studies, though even there my interests carry me a bit afield.
(Fitzpatrick, 2010)

Cluster A has absorbed various disciplines and developed as a vibrant interdisciplinary field. Although it has not torn down the walls between academic departments, it has made talking across those walls easier.

From the longitudinal network results (section 4.2.5 Five Periods), we can see that this cluster is a relatively recent part of the DH intellectual structure. It only appeared during the third period (1986-1990) with fewer than 50 cited scholars at the margin (Figure 4.7, on the right), and it quickly grew to a visible large cluster occupying around one third of the network nodes during 1991-2005 (Figure 4.8, at the bottom). During the last period (2006-2017), we can see that this cluster continued to grow and accounted for around half of the nodes on the network (Figure 4.9, on the left) with other clusters

moving further and further away from it. Such longitudinal networks show that cluster A only joined the field after 1990, and immediately attracted great attention from a variety of fields during the later times. This agrees with Hockey's observation that a number of new academic programmes were introduced from the early 1990s (Hockey, 2004, p. 13).

Nevertheless, cluster A only constitutes one part of the DH intellectual structure, although its subjects overlap with each other and have many similarities. While scholars in cluster A might seem more active and louder, the other 57.6% of scholars also made significant impact.

4.2.6.3 Cluster B – Computational linguistics

The cluster related to computational linguistics (or sometimes called contemporary corpus linguistics) is the second largest group on the network – cluster B (Figure 4.10). It is an interdisciplinary field which focuses on statistical modelling of natural language from a computational perspective and research on computational methods to linguistic questions (Nerbonne and Tonelli, 2016). Based on the current dataset, many cited scholars are related to this field, and most of their studies were published in the relatively earlier stage of DH development. For example, the book by Douglas Biber was published in 1988, and was cited the most in this cluster (43 times based on the current dataset) (Biber, 1988). Michael Halliday's book was cited 24 times (Halliday, 1985). It serves as an essential introduction to functional theory that has been widely used later in a number of applied linguistics contexts, particularly artificial intelligence and language education. The book written by Noam Chomsky in 1965, has 24 citations in the current dataset (Chomsky, 1965). It is widely considered to be the foundational reading of Chomsky's theoretical framework of linguistics – transformational-generative grammar.

Computational linguistics is the earliest area in DH (Jensen, 2014, p. 115). As can be seen from the longitudinal view of the author co-citation network, during the first period (1966-1970) in Figure 4.5, most of the cited authors were related to computational linguistics. For example, the most connected node, Joseph Kruskal, was an American mathematician, statistician, computer scientist and psychometrician. He applied his algorithm (i.e., Kruskal's Algorithm) in linguistics and conducted experimental

lexicostatistical studies of Indo-European languages that were cited the most during this period. During the second and third period (1971-1985, 1986-1990), cited scholars related to computational linguistics remained at the centre of the network, with a slight growth in number during 1971-1985 (Figure 4.6). While more and more scholars joined the whole network during 1986-1990 and later (Figure 4.7), the number of nodes in this cluster stayed approximately the same. Because it kept a stable number during the third period, it made space for the rise of cluster A (general historical literacy and information science) after 1990. While its dominant position was succeeded by cluster A during the fourth period (1991-2005) in Figure 4.8, cluster B (computational linguistics) itself was not replaced by the nodes in cluster A; instead, it moved away from the central stage during the last period (2006-2017) as shown in Figure 4.9.

Computational linguistics has not been mentioned very often in contemporary DH narratives (except, e.g., Hockey, 2004; Vanhoutte, 2013). This is probably because computational linguistics is often seen to have split off from DH and the practitioners have set up their own societies and conferences (e.g., Mitkov, 2014); apart from that, this section attempts to understand it from other different aspects.

Within this large cluster of computational linguistics (Figure 4.10), we can see many of the cited scholars are from a more ‘technical’ background, such as language modelling and natural language processing (e.g., David Yarowsky, Kenneth W Church). To some, computational linguistics seems to be more related to the quantitative (or practical) side of DH, instead of the qualitative (or interpretive) side of DH (Jensen, 2014, p. 128). The Digital Humanities Manifesto 2.0 used the metaphor of ‘waves’ to describe the development of DH (Schnapp and Presner, 2009), and Jensen argued that computational linguistics seemed to be in the first wave of DH (i.e., quantitative approaches) which was ‘washed away’ by the next wave (i.e., qualitative methods) (Jensen, 2014, p. 128). Jensen pointed to the possible reason that corpus linguistics (or computational linguistics in the context) was marginal in contemporary DH, which was related to its ‘quantitative nature’:

[Corpus linguistics]’s fringe status in DH is, perhaps, puzzling at first sight, but this is due to a number of complexities that reside in both the quantitative nature of CL with its focus on automation and the strict conception of contemporary DH as a qualitative and experiential second wave. It is clear

that, in the strict version of DH, CL in its entirety will forever be on the fringes because full inclusion would require CL to abandon its quantitative core. This is very unlikely to happen, given that CL is *per se* a framework within usage-based linguistics (Jensen, 2014, p. 131).

Jensen seemed to indicate that there was a strong relationship between its quantitative nature and the marginal status of computational linguistics in DH. Nevertheless, regardless of whether computational linguistics has only the 'quantitative' part in its nature (some argue it also has an apparent theoretical component (e.g., Uszkoreit, 2000)), the cluster B we see in Figure 4.10 does not seem marginal or on the fringe. It is the second largest cluster containing 30.3% of the nodes on the whole network, and it has even closer links to other DH clusters – cluster C (English studies) and cluster D (early pioneers) – than DH's largest cluster, A – general historical literacy and information science.

Computational linguistics, in fact, continues to develop and expand along with the evolution of digital technology, only this expansion has not been shown on the longitudinal co-citation networks. Leech discussed its expansion measurement based on the exponentially increased size of corpora, i.e., large bodies of computer-readable text:

Machine-readable text collections have grown from one million to almost a thousand million words in thirty years, so it would not be impossible to imagine a commensurate thousand-fold increase to one million million word corpora before 2021' (Leech, 2014, p. 10).

Those who work with computer corpora are suddenly finding themselves in an expanding universe. For years, corpus linguistics was the obsession of a small group which received little or no recognition from either linguistics or computer science. Now much is happening, and there is a demand for much more to happen in the future (Leech, 2014, p. 25).

In addition, Bowker found that there is a growing connection between contemporary corpus linguistics and library and information science (Bowker, 2018); because the former investigates samples of authentic language use, it can help the latter to 'better understand the literary warrant of a given text collection'. Indeed, many approaches in computational linguistics are quantitative and present a relatively high degree of

objectivity, but it also expects qualitative input to address questions in theoretical linguistics, cognitive science, and cognitive psychology. The qualitative part of computational linguistics requires theories of linguistic knowledge to interpret a person's need for generating and understanding language, and in such theories, areas like cognitive psychology play an important role in simulating linguistic competence (Uszkoreit, 2000). In other words, methods can be applied to conduct preliminary processing, but the results must be interpreted by the researcher.

The field of computational linguistics is clearly not shrinking, but why does it seem marginal according to the impression of the DH people? The reasons that some think computational linguistics is marginal in DH can be summarised as three factors. Firstly, there are few narratives in contemporary DH that focuses on the topic of computational linguistics, and especially in recent years; there are hardly any studies that reviewed the role of computational linguistics in DH, except, as mentioned, Jensen (Jensen, 2014). We, thus, might have the impression that there are fewer linguistic studies in DH. Secondly, some narratives seem to let us believe that we are currently at the second (i.e., qualitative and interpretative) wave of DH while the first (i.e., quantitative) wave has passed and gone (Schnapp and Presner, 2009); some, however, might not get this impression and see the point of a wave as being part of a sea where the waves are interconnected. As computational linguistics has been perceived by many as a technical area, it is often associated with the first wave and so seen in the past tense. However, as we discussed in section 1.1.3 ('Digital' and 'Humanities' question), 'quantitative' and 'qualitative' are not in binary opposition in DH, and they exist simultaneously and need to be taken care of in each DH project. Computational linguistics, on the other hand, is not solely about quantitative methods but also interpretive analysis from cognitive and psychological perspectives. Although the DH Manifesto 2.0 holds true to some extent, the first wave was not 'washed away', 'replaced', or 'vanished', but partly 'moved away' to some other place instead of staying in DH. Thirdly, as the technology development, corpus and text samples become increasingly large, computational linguistics has been linked very often to notions such as 'big data' and 'artificial intelligence' (Moreno and Redondo, 2016). Such a situation seems to associate the field to a more technical context, but a larger dataset actually generates more reliable and interpretable results that could support

qualitative analysis, despite the need for further selection and understanding when dealing with more results.

On the other hand, the cluster of computational linguistics in the overall visualised network (52-year period combined) is rather central, not marginal. Although we know from a longitudinal perspective that it was later pushed away and became smaller, the data accumulated during such a long period (52-year of the bibliometric dataset) might have evened out its recent marginal position and averaged its size. Computational linguistics in DH was indeed dominant in the past, as Jensen described in the first wave (Jensen, 2014), and it should be given more attention during the contemporary DH era rather than leaving it in the past or letting it move outside of DH.

In addition, new questions are emerging. This cluster has an apparent connection with the linguistic studies of non-English languages, especially German-Dutch linguistic studies (e.g., John Nerbonne, an American computational linguist who worked in Germany; Nerbonne's colleagues and co-authors at the University of Groningen, Peter Kleiweg and Wilbert Heeringa; Hans Goebel, an Austrian linguist). This question will be further addressed in section 5.2.6 (Discussion and analysis) and 5.3.6 (Discussion and analysis) to combine the results from Twitter analysis.

4.2.6.4 Cluster C – English studies

The subfield of English studies is the third largest cluster (cluster C) and contains around 19% of the cited scholars on the whole network. It is located in the centre of the network, indicating a high betweenness centrality and a strong bridging role to other subfields that reflects the breadth and reach of its contents. Most authors in this cluster are also able to reach to authors in the rest of the clusters.

As mentioned, compared to a small number of literature studies found in cluster A, studies in cluster C are more focused on analysing the literature texts as research subjects instead of analysing the field of literature studies. For example, there are many studies related to authorship, text analysis, and stylistics. The article (Burrows, 2002) by John Burrows was cited 45 times in the current dataset, which is the most cited publication in this cluster. This topic continued to become more and more popular when David L Hoover tested and commented on Burrows's method. The two articles by Hoover have been cited by 26 articles in the current study (Hoover, 2004a, 2004b).

David I Holmes, in this cluster, also had many publications of authorship studies that were cited, e.g., (Holmes and Forsyth, 1995) was cited 26 times, and (Holmes, 1994) was cited 18 times.

The works of Josephine Miles who, suggested by some, should 'replace' Busa as the founding figure of DH (Buurma and Heffernan, 2018), can also be found in this cluster, although the number of citations is not high (i.e., 6 times). Miles was an English professor at Berkeley, who started her first distant reading project in the 1930s. She was also the director of a concordance to *the Poetical Works of John Dryden* that was 'published 17 years before the first volumes of the 56-volume *Index Thomasticus* began to appear' (Buurma and Heffernan, 2018).

What is the relationship between DH and English studies? Many scholars think that the two are very closely connected. Kirschenbaum once raised the idea that 'digital humanities has accumulated a robust professional apparatus that is probably more rooted in English than any other departmental home' (Kirschenbaum, 2010, p. 3). Firstly, text-based data processing was and still is the dominant component of DH, and such method is helpful for research in authorship studies, stylistics, and linguistics that are mostly associated with language studies. This also explains why cluster C (English studies) has closer link to cluster B (computational linguistics) than cluster A (general historical literacy and information science), although the latter is also predominantly text-based, with more diverse topics seen in cluster A. Secondly, English departments, as Kirschenbaum argued, have long been hospitable and open to other areas (e.g., cultural studies using computational approaches) (Kirschenbaum, 2010, p. 9), and thus provide DH with a friendly environment in which to develop. Stutsman also supported this idea and found that 60% of the DH syllabi in her compiled dataset were based at English departments (Stutsman, 2013).

Nevertheless, according to many studies, it seems that one of the most heated debates in DH also emerged from English departments – the binary argument of the 'digital' and the 'humanities'. As mentioned earlier, it is a question of the balance between numbers and meaning (Liu, 2013, p. 411). As Pressman and Swanstrom once mentioned that literature studies were:

[...] a field dedicated to the interpretation and explication of meaning. Literary critics explain and interpret how meaning is made. They analyze texts in ways that illuminate the ideologies texts contain and thereby enable critique of them. Literary critics don't take data at its word. (Pressman and Swanstrom, 2013, para. 3)

This logic of the research behind literature studies appears to be contrary to the practical archetype of DH, although this is one that DH has pushed back against. Kirsch pointed out that the differences in research patterns between traditional literary practices of interpretation and DH practices of 'making' have caused tension, and 'technology is taking over English departments' because of 'the false promise of the digital humanities' (Kirsch, 2014). Pulizzi claimed that the advent of DH will make English departments pointless unless they 'pay more attention to the variety of media narrative' (Pulizzi, 2014). Risam reviewed that:

In such narratives, digital humanities has taken over English departments, reduced literature to mere 'data', and killed close reading. Much like 'theory' in the culture wars of the 1980s and 1990s, digital humanities is charged with contributing to the decline of the humanities. (Risam, 2018, p. 8)

These allegations make the impression that the field of English studies seemed to be worried because of the DH movement. Looking back at the visualised network (Figure 4.10), however, we hardly see any authors of such debates in cluster C (English studies). Authors of these debates are positioned within cluster A – general historical literacy and information science (e.g., Cecire, 2011; Sample, 2011; Liu, 2013, p. 411; Nowviskie, 2016a). Even the node of Matthew Kirschenbaum who authored the popular article 'What is digital humanities and what's it doing in English departments?' is positioned in the centre of cluster A, and his node has no co-citation link to cluster C at all. Therefore, based on the current dataset, we can see that most people who are doing English studies and related research in DH (e.g., authorship and stylistic studies) are hardly involved in the discussions about 'hack' and 'yack' or the future of English departments.

Nevertheless, we cannot say that scholars who are involved in the 'hack' or 'yack' debate in cluster A (general historical literacy and information science) are not from the field of English studies, because many of them are. For example, Matthew

Kirschenbaum is Professor of English in the Department of English at the University of Maryland, Stephen Ramsay is Associate University Professor of English at University of Nebraska–Lincoln, Alan Liu is Professor in the English Department at the University of California, Santa Barbara, and Bethany Nowviskie is Professor of English at James Madison University³⁰. There is no doubt that all of them are affiliated in the English departments, but their nodes are not placed within the cluster of English studies (i.e., cluster C). Scholars from one field are not always found in one cluster on the visualised network. It seems that the subject matter related to English studies (or the English department) in DH can be categorised mainly into two groups, English studies using digital methods and the discussion about English departments and DH. These two groups reside in cluster C and A, respectively, and the two clusters are not closely connected but loosely separated with a clear segregation.

With a close observation of the visualised network, we can find that English studies is one of the only two disciplines that have topics spread across different clusters on the network (the other discipline is history studies that has general historical literacy subject in cluster A and DH history subject in cluster D – early pioneers). It remains a question for future study as to whether this can be seen as an indication of subject integration within the realm of DH.

According to the longitudinal results, English studies appears to be important and central in the DH intellectual structure throughout the historical periods. As results in 4.2.4 Centrality have shown, the nodes related to the English cluster have higher betweenness centrality, indicating that these nodes have strong bridging roles to other clusters that reflect the breadth and reach of their subjects. It is also clear in Figure 4.10 that cluster C is in the centre of the whole network, and that it has a central and important role that connects to other subjects.

Similar to cluster A (general historical literacy and information science), cluster C joined the network rather late, after around 1985. It first appeared in the network during the third period (1986-1990) in Figure 4.7. Though small, its nodes (e.g., John Burrows)

³⁰ Although at the time of these debates, Bethany Nowviskie was in library and information science. She was the Executive Director of the Digital Library Federation (DLF) at CLIR, the Council on Library and Information Resources (2015-2019), and Director at Scholars' Lab and department of Digital Research & Scholarship, University of Virginia Library (2007-2015).

are located at the centre of the network, indicating an obvious bridging role to other areas from 1985 and onwards. During the next period – fourth period (1991-2005) in Figure 4.8 (on the right) – it has grown rapidly in size and accounted for roughly one third of the total node weight. From 1991, some cited scholars in this cluster (e.g., John Burrows, David Holmes) became top-cited in the whole network, doubling or tripling the sizes of other top-cited scholars; and John Burrows and David Holmes, in particular, remained the most-cited scholars in the whole period by the different counting methods (i.e., citation count, non-self citation count, fractional non-self citation count). The last period (2006-2017) in Figure 4.9 also witnessed the central position of English studies, but this cluster is clearly closer to the computational linguistic cluster, leaving a distinct gap from the cluster of general historical literacy and information science.

In addition, the central place of English studies is not difficult to understand as many studies have suggested this and implied that DH is ‘more rooted in English than any other departmental home’ (e.g., Kirschenbaum, 2010; Rice, 2013). Besides this, it is no news that the current ‘ADHO recognised’ DH community is dominated by the Anglophone (Galina, 2014). Broadly speaking, literary studies is the study, evaluation, and interpretation of literature in any language (Groden et al., 2005), but in the context of DH, it is mainly limited to English literary studies, and sometimes ‘English studies’ and ‘literary studies’ are used interchangeably. The wider environment also encourages such use as English became the language of the Internet and the lingua franca of the Web, and global academic systems are in favour of English publications (Fiormonte, 2014; Mahony, 2018). Therefore, it is not surprising to find many English speakers in DH and many DH scholars who study some aspect of English literature.

However, the dominant English-speaking situation also has a positive side. Some scholars have found that the Anglophone DH community is very open to initiatives in other languages and keen to foster a more international landscape (Pitman and Taylor, 2017). While the Anglophone community still has a long way to go to achieve a ‘global DH’, the future is believed to be positive (Mahony, 2018).

4.2.6.5 Cluster D – Early pioneers

This cluster that contains many DH early pioneers is the last main group on the visualised network – cluster D. It is distantly connected to the other clusters, and most nodes in cluster D do not link to the rest of the nodes, except for some loose connections to cluster C (English studies). Although it is relatively small, it still has around 300 cited scholars, and many of them are recognised as the pioneers of DH, e.g., Roberto Busa (cited 45 times), Andrew Morton (59), Tom Merriam (75), Roy Wisbey (27), and Wilhelm Ott (21).

Despite the early contributions they made to the field, there is another force that strongly pulls these co-cited nodes of pioneers together as a cluster – the urgent need for DH history. As mentioned earlier, the history studies of DH have only started to emerge recently and are considered by some to be absolutely necessary (Nyhan and Flinn, 2016, p. 14). However, according to the longitudinal networks, these pioneers had already been grouped into an obvious cluster since the second period (1971-1985) in Figure 4.6, although the size of the cluster continued to shrink during later periods. It appears that there are already a number of studies published during 1971-1985 referring to them and citing their works together. Nevertheless, we should note that making citations of pioneers' works is usually for building on with new knowledge in a particular domain, while studying the history of the field is typically to look at examples in the longer patterns so that one can detect the trajectory of research development. These two activities are different.

Looking closely at the dataset, we can see that the well-known list of DH pioneers that we know today started to emerge around the time of a series of publications by Dolores M. Burton. In 1981, she published the first article in *CHum* titled 'Automated concordances and word indexes: The fifties' which reviewed and discussed the concordances between 1950 to 1959 (Burton, 1981a). In this article, Burton referred to many pioneers and cited their works, e.g., Roberto Busa, Stephen Parrish, Roy Wisbey, Antonio Zampolli. In the same year, her second article described the history in the early 1960s and the establishment of early DH centres, especially how Roy Wisbey founded the Literary and Linguistic Computing Centre at the University of Cambridge, the contributions of Andrew Morton, and a few other scholars including Busa (Burton, 1981b). Immediately after, her third article described the related projects

in the late 1960s and 1970s (Burton, 1981c). Similarly, this article cited many scholars that we know today, e.g., Stephen Parrish, Andrew Morton, Roy Wisbey, Roberto Busa, Wilhelm Ott, Trevor Howard-Hill, and Susan Hockey, although there are many other names, too. Burton's final article included more critical discussions about quantitative versus qualitative instead of listing contributions, nevertheless, familiar names were still mentioned. All these names mentioned above also appeared in the cluster D although their number of citations vary.

Burton's publications provided a great deal of historical detail and narratives about the early history of DH, and laid the foundation for later DH history studies, e.g., Hockey (Hockey, 2004) and McCarty (McCarty, 2003a). However, empirical studies showed that citations can be biased, especially when it comes to secondary source citations (i.e., citing a secondary source that discusses and cites information originally presented in a primary source) (MacRoberts and MacRoberts, 1996, p. 436). As there are only a handful of DH history studies, scholars who read them are more likely to cite what those studies have cited. Moreover, these cited pioneers also tended to cite each other forming a 'small world' of well-known DH pioneers, although it might also be the case that they were building on one another's work. For example, Busa mentioned Antonio Zampolli (who was his assistant and later founded the laboratory for computational linguistics in Pisa) very often in his publications (e.g., Busa, 1980, p. 86, 2004, p. xvi). Roy Wisbey started to mention Busa's projects in his articles as early as in 1962 (Wisbey, 1962). Thus, later studies are more likely to cite the same group of people after reading these publications.

Consequently, we may ask if the early history of DH is really a homogenous one? Although it seems to be this particular group of people doing DH-related work at the beginning, many studies think otherwise. For example, many studies consider that the DH history is heterogeneous and question the founding father position of Busa (McCarty, 2003b; Nyhan and Flinn, 2016; Jones et al., 2017). Some scholars have suggested replacing Busa with Josephine Miles (Buurma and Heffernan, 2018), or even to trace the earliest DH scholar to Thomas Mendenhall in the 1880s (Pinsken, 2009), and the two scholars' names can also be found in the visualised network.

Sula and Hill pointed to the four concerns brought about by homogenous history studies in DH

Though this account [previous narrative of DH history] dominates historical views of the field, it raises four separate concerns. First, it privileges certain disciplines, projects, and tools at the expense of others (e.g., quantitative history, which is absent from the narrative). Second, it fails to chart an actual historical path from early work in text analysis to ‘big tent’ DH, encompassing everything from digital archives and databases to GIS, network analysis, new publishing formats, digital pedagogy, and so on. Third, it precludes historicizing and contextualizing current work that falls outside of text analysis, which may lead to a lack of attention to method, its historical complexities, and points of convergence with related fields such as the social sciences. Finally, these histories all suffer from a lack of evidence; the narrative is assumed and applied rather than documented. (Sula and Hill, 2017)

Many believe that the lack of heterogeneous history and a more comprehensive perspective to study it prevents DH’s disciplinary development and future improvement (McCarty, 2011, pp. 4–6; Nyhan and Flinn, 2016, p. 15), and McCarty even described it as a ‘crying need for history’ (McCarty, 2011, p. 6). By studying the heterogeneous history of DH, one can understand how many other DH scholars have reached the current stage and what historical changes have shifted this field from its past to its present position.

To study such a history, one must look back at the earliest figures in the field and their works. By reviewing the nodes presented in this cluster and early nodes in other clusters, we can gain more insights into the DH early history than hand-picking the well-known scholars from previous narratives. Although we could trace the field’s earliest history to Roberto Busa in the 1950s (Busa, 1950) or Josephine Miles in the 1930s (Buurma and Heffernan, 2018), many other lesser-known figures are also shown in the results, and the history can seemingly be dated back to the nineteenth century or even earlier.

For instance, by looking at the publication year of the cited references in the current dataset, we can learn that before Roberto Busa started the work of *Index Thomisticus* in 1949, many similar works were also conducted. Although the collected articles in the current dataset only dated back to 1966, and thus could not be visualised as networks to show the period before then (not to mention the period before 1949), their cited references can make up for this deficiency.

Among all the 49,047 references that were cited by the collected articles, 7,420 of them (15.13%) were originally published before 1949. Why did DH scholars cite so many works that were published before the time of Busa? If we take a look at the most cited authors who published works before 1949, things will become clearer (see below Figure 4.11). There are generally two reasons for DH scholars to cite works published before 1949. One is that the work itself was the subject of study, for example, William Shakespeare was cited 20 times by the articles in the current dataset. The texts written by Shakespeare have been analysed by many DH scholars, but Shakespeare himself was not considered a DH scholar. There are many highly cited authors in the current dataset who published works before 1949 that belong to this group, e.g., Joseph Conrad and Virginia Woolf. This will be discussed with more examples later. Another reason for this is to refer to the early relevant methods, techniques, and projects that were conducted by early scholars in this field and other related fields that DH was in conversation with, e.g., Alan Turing, Thomas Mendenhall, Sigmund Freud.

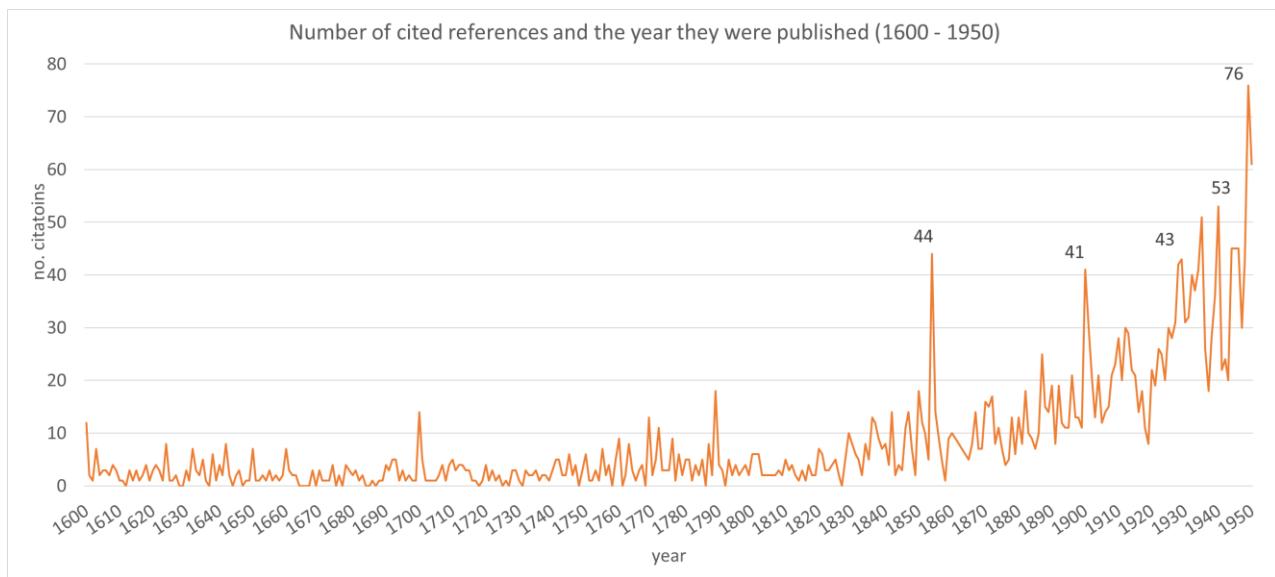


Figure 4.11: The number of cited references that were originally published between 1600 to 1950 in the current dataset.

Figure 4.11 above shows the number of cited references that were originally published from 1600 to 1950 each year in the current dataset. The number fluctuates over time, and there are several peaks above the line of 20. Some of these peaks, especially the ones before 1880, were due to the first reason mentioned above. For example, there is a peak with 44 cited references to 1854. Why did articles from *CHum*, *LLC/DSH*,

and *DHQ* cite 44 references published in 1854? This is because a few studies researched events in 1854. One of them was done by Wolff who analysed 396 roll-calls that led to the 1854 Kansas-Nebraska Bill using IBM cards and computational methods, and in this study he cited 20 archives of roll-calls that were published in 1854 (Wolff, 1974).

On the other hand, many peaks between 1880 to 1950 in Figure 4.11 represent the number of citations received by some pre-Busa scholars who might be from other related fields, and whom we do not know well or are not familiar with. These numbers also indicate the influences they made to the field that have not yet been acknowledged extensively in the DH community. Table 4.3 below shows the top 60 authors whose works were published during 1880-1950 and cited the most. Authors that were cited for the two reasons mentioned above are both listed in the table. Although it requires further effort and analysis to distinguish the scholars who indeed made contributions to the early field formation from the authors whose works were cited because others were doing analysis on their texts, one can already see some names as examples. This study chooses to keep both groups of names in the list, instead of removing the authors whose works were being studied as subjects, in order to keep the data as complete as possible for future analysis.

Table 4.3: The top ranked 60 authors whose works published during 1880-1950 were cited the most.

	name	citations		name	citations	
1	George Yule	35		31	Janet AH Murray	5
2	George Zipf	16		32	John A Scott	5
3	Willa Cather	15		33	Edward H Simpson	5
4	Thomas Mendenhall	14		34	Alan M Turing	5
5	Leonard Bloomfield	13		35	Carrington B Williams	5
6	Vannevar Bush	10		36	S Anderson	4
7	Sigmund Freud	10		37	Frederic C Bartlett	4
8	Claude Shannon	10		38	Bertrand H Bronson	4
9	Raymond Chandler	9		39	Joseph Conrad	4
10	Walter Willson Cobbett	8		40	Arthur C Doyle	4
11	GP Krapp	8		41	William Faulkner	4
12	Hans Kurath	8		42	Gottlob Frege	4
13	Ronald Crane	7		43	JA Hawkins	4
14	Ronald Fisher	7		44	Charles J Kappler	4
15	Arthur Friedman	7		45	JM Manly	4
16	Ivor Richards	7		46	Conrad Mascol	4
17	Edward Sapir	7		47	T Mommsen	4
18	S Chambers	6		48	AC Pickett	4
19	Thomas Eliot	6		49	John M Robertson	4
20	Walter Greg	6		50	Burrhus F Skinner	4
21	Alfred Housman	6		51	Virginia Woolf	4
22	D Ogden	6		52	DC Allen	3
23	Beatrix Potter	6		53	Aristotle	3
24	Vladimir Propp	6		54	L Frank Baum	3
25	HF Smith	6		55	V Berard	3
26	CF Brooke	5		56	M Bloch	3
27	F Brown	5		57	F Boas	3
28	CM Firth	5		58	DP Boder	3
29	Henry James	5		59	Roberto Busa	3
30	Otto Jespersen	5		60	Alonzo Church	3

As can be seen from Table 4.3, the most cited authors who published their works between 1880 and 1950 were not very well mentioned by scholars who studied DH history (as far as the candidate is aware). One cannot find many early pioneers on this list, for example, there is no Wilhelm Ott, Andrew Morton, or Roy Wisbey, and even Busa is ranked 59th with 3 citations. This is because most of Busa's works were

published after 1950. He only got very few publications before 1950, e.g., (Busa, 1948, 1949), so did other well-known DH pioneers.

What were the main topics (or subject specialties) that formed this field before Busa (if there was a field)? As Table 4.3 shows, apart from the authors whose writings were studied as the research topics (e.g., Willa Cather³¹), there are statisticians and mathematicians (e.g., George Yule³², Claude Shannon³³), linguists (e.g., George Zipf³⁴, Leonard Bloomfield³⁵), physicists (e.g., Thomas Mendenhall³⁶), engineers (e.g., Vannevar Bush³⁷), neurologists (e.g., Sigmund Freud³⁸) and many others who indirectly contributed to the early knowledge and formation of the field or contributed to part of a pool of knowledge that the field of DH was in conversation with. Unlike what we think of as early quantitative studies in the humanities, these authors are from a diverse and dynamic range of disciplines, and are not limited to Classics, linguistics, and literary studies. Moreover, Table 4.3 provides a starting point for people to study these early contributors in the future. For example, it is worth researching from the DH history perspective to establish how and why these authors were being cited in these journal articles and giving concrete examples about how their works were taken up by the DH communities.

Although much analysis is required in the future, this section begins that task by giving general introductions to some of the authors. Many achievements made by these authors before 1949 not only inspired the origin of the field that we call digital humanities today but also laid the technical foundations of it. For example, '*The Statistical Study of Literary Vocabulary*' by Yule (who is ranked the first in Table 4.3)

³¹ Willa Cather was an American writer, her novels includes *O Pioneers!* (1913), *The Song of the Lark* (1915), and *My Ántonia* (1918).

³² George Yule, usually known as Udny Yule, was a British statistician.

³³ Claude Shannon, an American mathematician, electrical engineer, and cryptographer known as 'the father of information theory'.

³⁴ George Zipf, was an American linguist and philologist who studied statistical occurrences in different languages.

³⁵ Leonard Bloomfield, an American linguist, whose linguistic approach is characterised by an emphasis on the scientific basis of linguistics and the formal process of linguistic data analysis.

³⁶ Thomas Mendenhall, an American autodidact physicist and meteorologist.

³⁷ Vannevar Bush, an American engineer and inventor who is known particularly for his engineering work on analog computers and the memex.

³⁸ Sigmund Freud was an Austrian neurologist and the founder of psychoanalysis, a clinical method through dialogue text analysis between a patient and a psychoanalyst.

was one of the earliest efforts of literary detection using statistical methods (Krippendorff, 2018, p. 108). Zipf (the second in Table 4.3) established and popularised the Zipf's Law in 1935 which was one of the first academic studies of word frequency, and is still used widely (Kanwal et al., 2017; Wang et al., 2017). Bloomfield (the 5th in Table 4.3) influenced the development of structural linguistics and helped to form linguistics as an established and defined science. Although the structural linguistics of Bloomfield was later criticised by Noam Chomsky (who is sometimes called 'the father of modern linguistics') and was overshadowed by Chomsky's generative grammar theory, their research focuses have much in common and are often compared and discussed together (Koerner, 1989). The memex and publication '*As We May Think*' of Bush (the 6th in Table 4.3) influenced generations of computer scientists, who drew inspiration from his vision of the future. Shannon (the 8th in Table 4.3), Bush's graduate student, known as 'the father of information theory', published '*A Mathematical Theory of Communication*' in 1948 which was believed to be a landmark that founded information theory.

These achievements were accomplished before the arrival of modern computers. It helps to show that the quantitative approaches to the study of the humanities are not something led by DH as its own innovations, just as Nyhan and Flinn suggested (Nyhan and Flinn, 2016, p. 2), but it might be these accumulated practices and works that led to the emergence of the field.

As DH was not (and perhaps now has still been developing into) an integrated field before and during the time of Busa, it is difficult to identify scholars who were doing DH-related studies, and they could be in any of the above-mentioned disciplines. Going through each related field, researching their previous scholars, and finding DH scholars is a job that requires a large amount of labour and time. The names in Figure 4.11 and Table 4.3, therefore, provide a new perspective to revisit the history of DH for future studies.

4.3 Co-authorship network

The co-authorship network is the second of the four networks that this thesis constructs. Co-authorship among scholars is used as a proxy for research collaborations to measure social connections of academic communities (Kumar, 2015,

p. 55). As can be seen from the current dataset, the patterns of scientific communication and collaboration in DH have been gradually developing and changing overtime.

Firstly, from a descriptive statistical point of view, there are a total of 2,527 articles, and 960 of them are multi-authored articles, which accounts for 37.99% of the total number. Figure 4.12 and Figure 4.13 present the proportion and percentage of co-authored articles over the full 52-year period. There is a steady growth on both graphs, and the multi-authored articles have accounted for the majority, especially after 2006, e.g., *LLC/DSH* in 2013 (69.49%), 2015 (66.07%), 2017 (70.24%). Although the level of co-authored articles does not fully represent the pattern of collaboration, it is one of the important indicators and can help us gain more understanding of the DH academic collaboration.

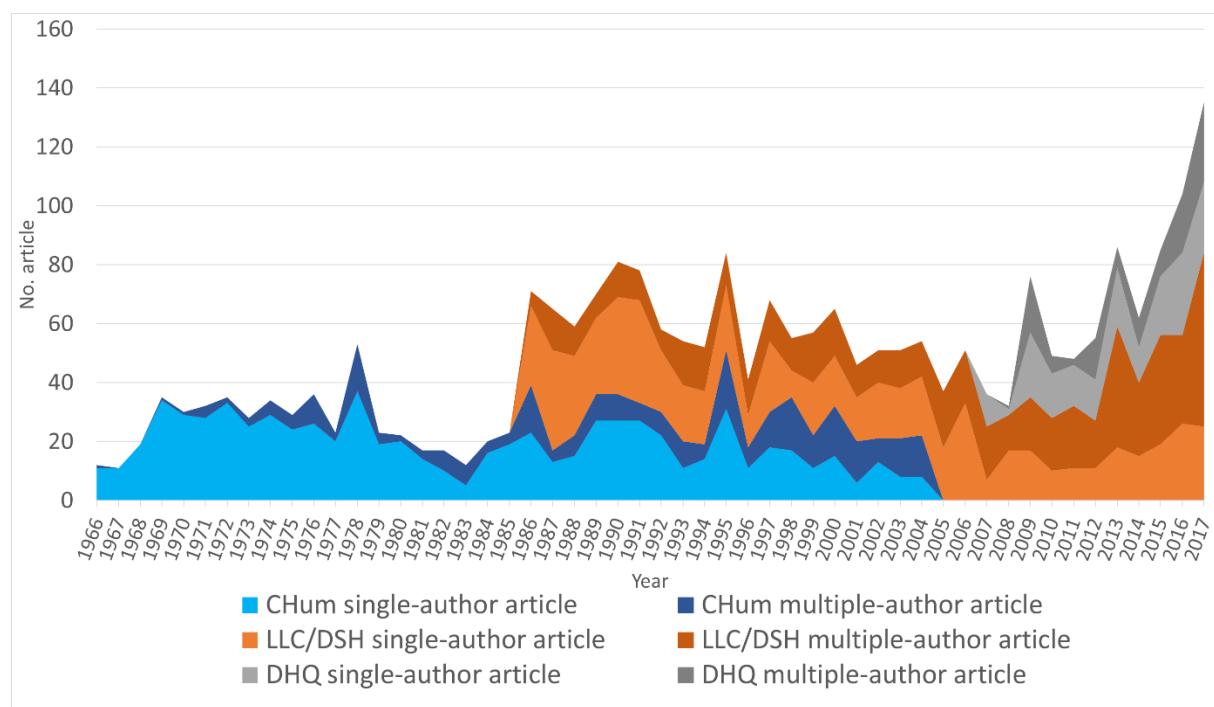


Figure 4.12: The number of single-authored and multiple-authored articles collected/published each year in the journals *CHum*, *LLC/DSH*, and *DHQ* (1966-2017).

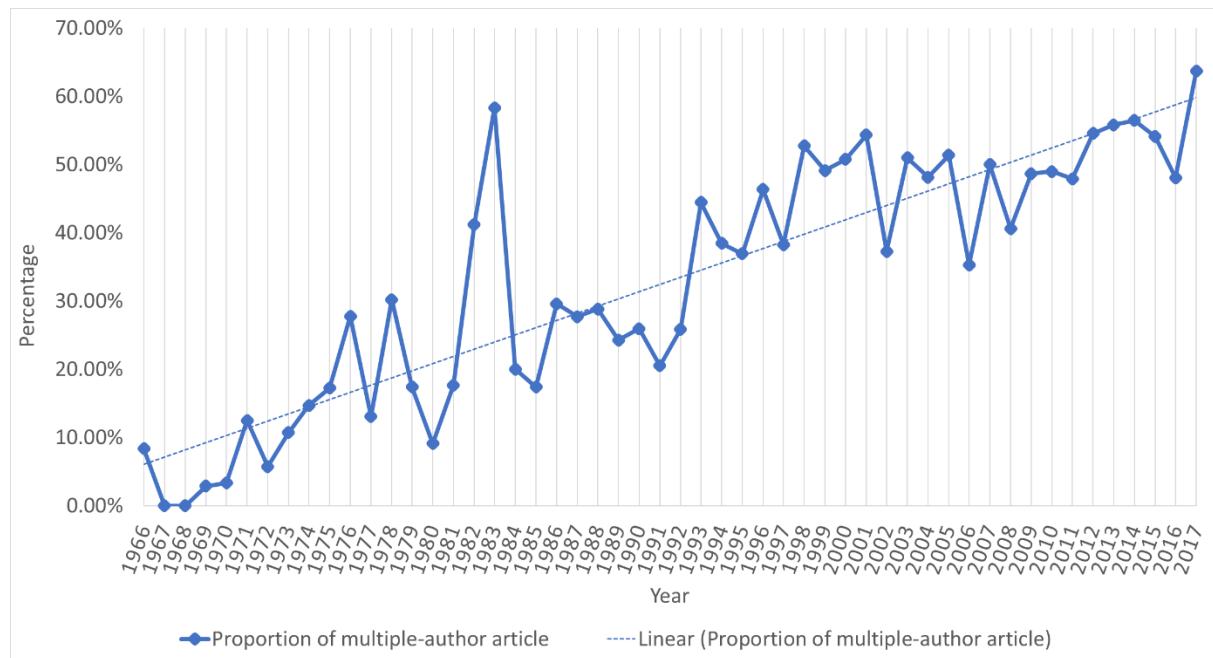


Figure 4.13: The relative percentage of multiple-authored articles each year in the collected dataset from the journals *CHum*, *LLC/DSH*, and *DHQ* (1966-2017).

However, perhaps the increase in co-authorship is not something unique in DH (e.g., Nyhan and Duke-Williams, 2014b). In order to explore the research question about who has contributed to the development of DH and how, this study goes beyond descriptive statistical analysis and employs network method to visualise the social links of the DH communities.

As mentioned in chapter 3 (3.3.1 Co-authorship network), this study counts the total number of publications to weigh the author node (section 4.3.1) and the number of co-authored publications for the edge (section 4.3.2). VOSviewer and Gephi have also been used for network visualisation and centrality measures, and the longitudinal periods were the same as in the ACA study.

4.3.1 Node

There are in total 4,623 authors who have published in *CHum*, *LLC/DSH*, and *DHQ* up until December 2017, and 3,382 unique authors have been identified. The mean number of authors per article is 1.83, and each author has published 0.75 articles on average (in total of 2,527 articles in the dataset).

The node has been weighted by counting the total number of articles each author published in the dataset (i.e., full publication count). Although the total number of publications cannot directly link to the author's academic productivity (Rørstad and Aksnes, 2015, p. 318), especially from a limited number of journals, it still reflects the contributions of the author within the data range of this study.

Table 4.4 below shows the top 50 authors ranked by the number of publications in these journals, and the complete table can be found in Appendix B. The author names have been used as node labels, and the size of the node has been weighted by its number of articles.

Table 4.4: The top 50 authors by the number of publications in journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

	Author Name	No. Articles		Author Name	No. Articles
1	Melissa Terras	27	26	Geoffrey Rockwell	8
2	Susan Hockey	22	27	Michael Sperberg-McQueen	8
3	MWA Smith	20	28	Tony McEnery	8
4	Mark Olsen	17	29	Gregory Crane	8
5	John Nerbonne	16	30	R L Widmann	8
6	David Holmes	16	31	Arianna Ciula	8
7	Thomas N. Corns	16	32	William Kretzschmar	8
8	Edward Vanhoutte	15	33	Susan Schreibman	8
9	Willard McCarty	14	34	Nancy Ide	8
10	Peter Robinson	14	35	G Lessard	8
11	Julianne Nyhan	13	36	Ellen Johnson	8
12	Estelle Irizarry	13	37	Matthew Spencer	8
13	Susan Brown	12	38	WJ Jones	8
14	Fiona Tweedie	12	39	Anne Welsh	7
15	Claire Warwick	11	40	Harold Short	7
16	John Bradley	11	41	Shlomo Argamon	7
17	Paul Fortier	11	42	Patrick Juola	7
18	Lou Burnard	11	43	Espen Ore	7
19	Christopher Howe	10	44	Richard Frautschi	7
20	Lisa Lena Opas-Hänninen	10	45	Antonio Zampolli	6
21	Stan Ruecker	9	46	Richard Forsyth	6
22	Julia Flanders	9	47	JR Allen	6
23	Whitney Bolton	9	48	Michael Levison	6
24	Barron Brainerd	9	49	T Rommel	6
25	Raymond Siemens	8	50	SG Hall	5

4.3.2 Edge

To calculate the edges of the co-authorship is straight forward. Each edge represents a co-authorship link, and if any two authors co-authored an article, then the value of their co-authorship edge increases by 1. The undirected edges of each pair of co-authors were calculated and the matrix structure is similar to that of the ACA network (section 4.2.2). The complete network data can be found in Appendix F.

4.3.3 Network visualisation

Among the 3,382 identified authors, only 661 (19.54%) were connected to the main network of co-authors. John Burrows, who is the most cited scholar in the dataset, published 10 articles and co-authored one paper (with Hugh Craig in *CHum*) (Burrows and Craig, 1994), but he was not able to connect to the main co-authorship network as he has no link to any node on the main network. Figure 4.14 below shows the co-authorship network with all the authors in the dataset.



Figure 4.14: The co-authorship network with all the authors.

According to Figure 4.14, the main network at the centre accounts for 19.54% of the nodes, and there are other smaller networks and individual authors that are detached from the main network and scattered around. Most of the authors who published more

than one article have been connected to the main network, although some have formed smaller networks that are disconnected from the main one, e.g., John Burrows, Gordon Dixon, Robert J. Valenza, Roberto Busa. In general, however, the more articles an author has published in these journals the more likely that this author is able to connect to the main network. Figure 4.15 below shows the main co-authorship network.

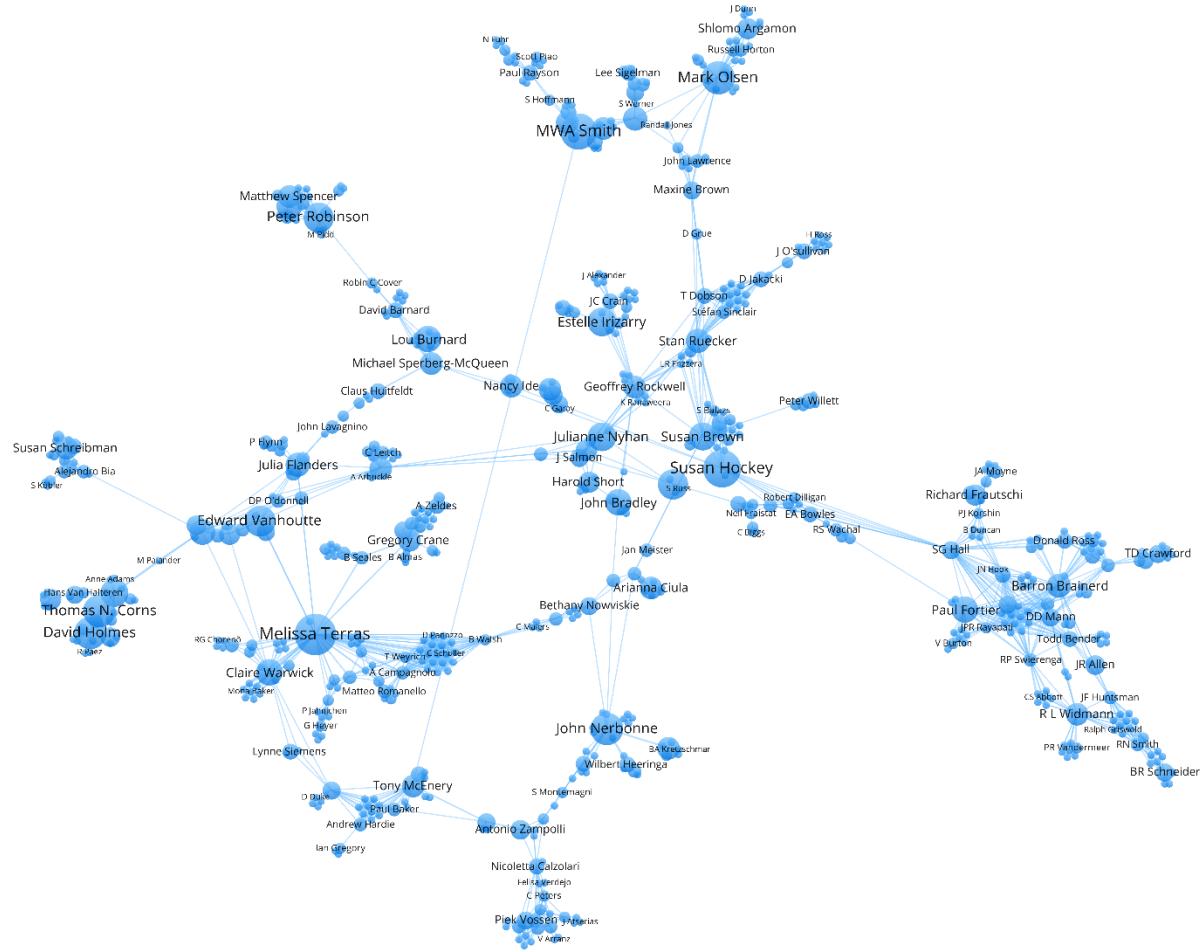


Figure 4.15: DH co-authorship network, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017), graph created by VOSviewer

4.3.4 Centrality

As shown in Figure 4.15, it is clear that some nodes are important bridges of the network, such as Susan Hockey, Julianne Nyhan, Melissa Terras, Edward Vanhoutte, and John Nerbonne. To better investigate the roles that scholars play in research collaboration, it is beneficial to apply centrality measures. Studies have shown promising advantages for the use of centrality indices to analyse and interpret co-

authorship networks (De Stefano et al., 2011; Abbasi et al., 2012). As mentioned earlier (section 4.2.4 Centrality), this study has employed Gephi 0.9.2 to calculate the *betweenness centrality* for all the four networks and the process is similar to that of the ACA study in section 4.2 (ACA network). Table 4.5 below shows the top 10 authors who have the highest values of betweenness centrality; the complete table can be found in Appendix F. These values reflect how important a node's position is on the network.

Table 4.5: The top 10 authors by the value of Betweenness Centrality on the co-authorship network, data extracted from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

	Author Name	No. co-authors	No. publication	Betweenness Centrality
1	Susan Hockey	16	22	81501.32
2	Susan Brown	25	12	68457.91
3	SG Hall	36	5	67345.25
4	Raymond Siemens	18	8	57718
5	Melissa Terras	53	27	49180.26
6	Stan Ruecker	31	9	48101.25
7	Julianne Nyhan	13	13	36533.59
8	Geoffrey Rockwell	23	8	34623.23
9	Maxine Brown	9	5	31533.05
10	Paul Spence	7	3	30916.79

4.3.5 Longitudinal analysis

As reviewed in chapter 2 Literature Review (section 2.2.1 Co-authorship), the proportion of authors that are connected to the main co-authorship network is very limited. DH, in particular, often forms as 'small world' models with a small number of authors. The scale of the co-authorship network depends heavily on the scale of the dataset, i.e., the number of publications. In other words, the more publications one collects, the more authors who are connected to the main co-authorship network. Because of the low proportion of authors in the main network, when studied longitudinally, the dataset can no longer be sliced into separate periods as in the ACA study (section 4.2, ACA network), but as an accumulated structure as a whole. If we

look at the following Table 4.6 of the percentages of the connected authors during each separate period, we can see that the number of connected nodes is very low.

Table 4.6: The relative percentages of the connected authors during different period on the co-authorship networks, data extracted from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

Time period	No. authors	No. connected authors	Percentage of connected authors
1966-1970	72	3	4.17%
1971-1985	383	6	1.57%
1986-1990	381	6	1.57%
1991-2005	1,076	52	4.83%
2006-2017	1,440	96	6.67%
whole period (1955-2017)	3,382	661	19.54%

As the very low numbers of connected authors within each separate period cannot provide useful information to study the pattern of scholarly collaboration, this study has combined the periods and constructed an accumulated co-authorship network. By using different colours to code the average publication year of each author, Figure 4.16 below shows the accumulated longitudinal information for each co-author.

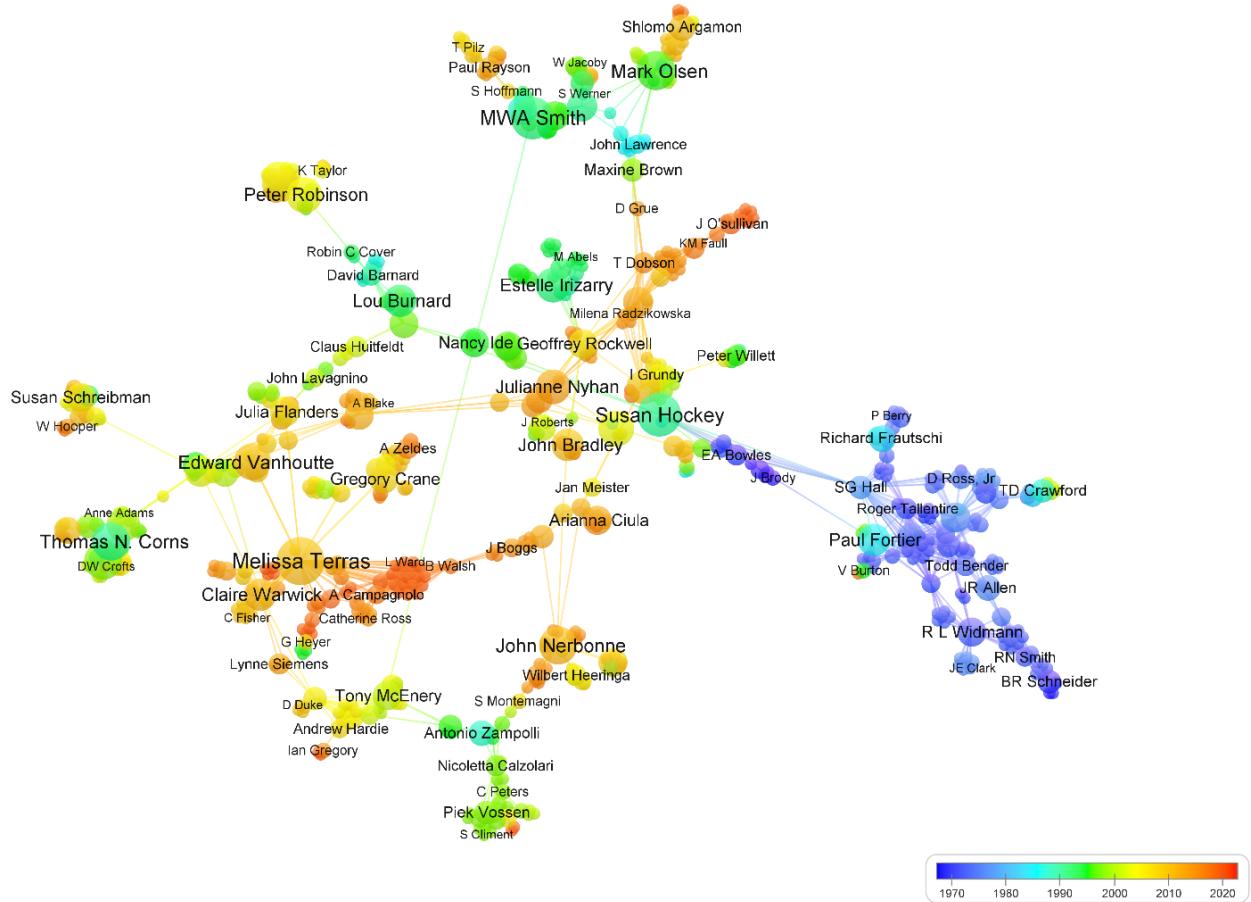


Figure 4.16: DH co-authorship network with average year information, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017), graph created by VOSviewer

With the colour-coding in Figure 4.16, we can see the different positions that authors belong to at different times, and it reflects the collaboration trend of authors overtime. Especially on the right-hand side, we can see that people who published articles before 1980 tend to be linked together. Nevertheless, we can also find some authors who are consecutively highly ranked in all time periods, e.g., Susan Hockey, indicating their ‘career plateau’ throughout the time (Chang Boon Lee, 2003, p. 538). Comparatively, some authors are on the rise in this field (in orange and red) while some are faded out (in blue).

4.3.6 Diversity networks

Adding different dimensions into the network helps to explore the formation and influencers of the network (Badar et al., 2013, pp. 767–773). As mentioned, this study has chosen gender and affiliated country as the two dimensions to assist studying the

collaboration pattern in DH. To avoid duplicate network analysis and interpretation, results and discussion will be presented at the end of this section (4.3.7 Discussion and analysis).

4.3.7 Discussion and analysis

This section addresses the *scholar* and *environment* research questions within the bibliometric context from three perspectives - 4.3.7.1 Historical periods, 4.3.7.2 Gender, and 4.3.7.3 Country. Each perspective represents a factor that this study considers might influence the DH co-authorship and the formation of its collaborative community.

As shown in the co-authorship results, we can see that the collaborative nature of DH was not found during the early days of this field, or at least such nature was not shown in publications. It has, nevertheless, been developing throughout the expansion of DH and is becoming increasingly notable especially in the past decade (4.3.7.1 Historical periods). Environmental factors (e.g., gender, language, and affiliated country) have played important parts in the formation of DH scholars' co-authorship network. Although male scholars have dominated the field, female scholars have experienced a rapid growth during the last 20 years, and they have been acting as critical bridges in forming collaborative links (4.3.7.2 Gender). Authors affiliated in Anglophone countries are the majority in the author pool. While the US scholars accounted for a large proportion of the publications, scholars in the UK and Canada were more likely to contribute to the formation of co-authorship links than any other country both as bridges as well as central nodes. On the other hand, it is interesting to notice that the level of international collaboration in DH is more extensive than many other disciplines, indicating an ever-growing international collaborative community (4.3.7.3 Country).

4.3.7.1 Historical periods

Based on the results and previous studies, there is a correlation between the level of co-authorship in DH and the historical periods. It seems that the more recent the publication, the more likely that it is co-authored rather than single-authored.

In the current dataset, only 37.99% of articles were co-authored (*CHum* 27.92%, *LLC/DSH* 46.53%, *DHQ* 38.72%). This figure challenges the collaborative character

that is usually depicted for DH, because it is lower than many results from previous DH co-authorship studies. As mentioned in chapter 2 (Literature Review section 2.2.1), Spiro found that 48.28% of *LLC* articles (2004 – 2008) were multi-authored (Spiro, 2009), Weingart reported that the multi-authored papers at DH2013 – 2016 stayed at around 60% (Weingart, 2013a, 2013a, 2014a, 2015a, 2016b). The Dutch-speaking DH conference DHBenelux had 58%, 60% and 66% of multi-authored papers from 2016 to 2018 (Kemman, 2016b, 2017, 2018), and the German DH conference (DHd2018) had 72% of multi-authored papers (Henny-Krahmer and Sahle, 2018).

The reason that this study has 37.99%, which is lower than many previous figures, is probably due to the different time spans that these studies selected to compile their datasets. If we investigate these different percentages further, we find that the more recent the time span the higher the percentage. For example, during 2004 – 2008 the percentage was 48.28% in Spiro's study (Spiro, 2009), while in 2018, it was 66% and 72% in two studies (Henny-Krahmer and Sahle, 2018; Kemman, 2018). Whereas, the current study has the percentage as 37.99% because it covers a wider time range (1966 – 2017), and the 1966 – 2011 dataset collected by Nyhan and Duke-Williams had even lower figures (31% *CHum*, 35% *LLC*) as it did not cover more recent publications (i.e., publications after 2011) (Nyhan and Duke-Williams, 2014a, p. 392). In other words, based on these figures, DH scholars tended to publish single-authored articles earlier while the proportion of co-authorship has been gradually increasing overtime. Figure 4.12 and Figure 4.13 were presented earlier in this section, and they are referenced below with the same figure numbers to demonstrate the number and percentage of multi-authored articles in the current dataset.

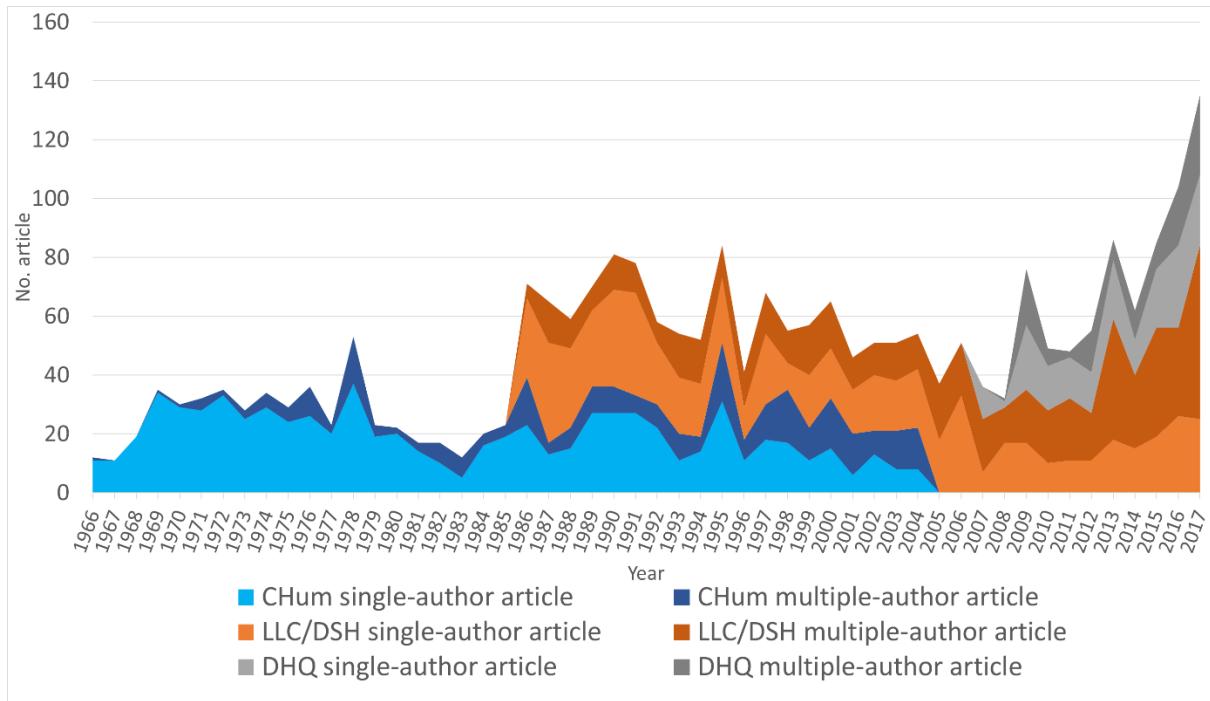


Figure 4.12: The number of single-authored and multiple-authored articles collected/published each year in the journals *CHum*, *LLC/DSH*, and *DHQ* (1966-2017).

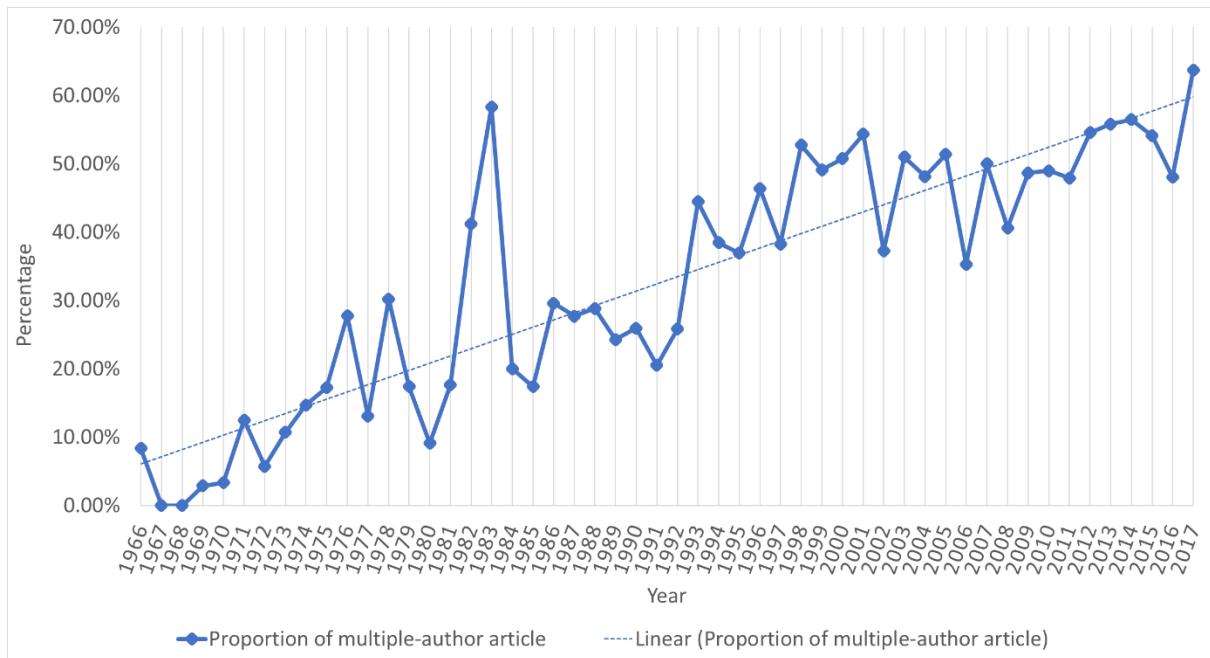


Figure 4.13: The relative percentage of multiple-authored articles each year in the collected dataset from the journals *CHum*, *LLC/DSH*, and *DHQ* (1966-2017).

In general, the percentage of co-authored articles in the current dataset has been increasing steadily with an upward trend, despite the spike in 1983 which was an

anomaly due to the general lack of publications (see Figure 4.12) in that year. Especially in more recent years, the proportion of co-authored articles exceeded that of the single-authored ones, e.g., co-authored articles account for 69.49% (2013), 66.07% (2015), and 70.24% (2017) in *LLC/DSH*. These numbers, however, are not contradictory to some earlier studies that showed that DH was dominated by single-author publishing patterns (Nyhan and Duke-Williams, 2014a, p. 387; De la Cruz et al., 2015, p. 4), because these studies were conducted in different time periods.

Why do we see more and more co-authored publications in DH? From a broader environmental view, academic collaboration in general has been changing, and it brings different academic communities from around the globe closer. DH is not unique in the growth of scholarly collaboration, and studies have shown that the co-authorship increase has been seen in almost all disciplines since the post-World War II period (Cronin, 2005; Wuchty et al., 2007). Although disciplines like Sciences, Engineering, and Social Sciences have more significant growth in co-authorship, the Humanities also has observed an increase in multi-authored publications (Wuchty et al., 2007, p. 1037). Studies have shown co-authorship growth in DH related fields, such as Library Studies (Ding, 2011; Cheng et al., 2019), English Studies (Leane et al., 2019), and Music Studies (Layman and Elliott, 2019). In other words, because the publishing environment is changing and the whole of academia is having more co-authored publications over time, it is not surprising to see that DH similarly has more and more co-authored papers.

The external reasons for such an increase are various, and some are apparent. Apart from the improvement of communication channels, the policy-making preference, the funding of groups of scholars instead of funding individuals has become more 'formalised' for its higher chances of effectiveness and productivity (Wuchty et al., 2007). Groups of scholars are more likely to bring diverse specialties, and teamwork has been proven to exceed and increasingly dominate individuals during knowledge production (Kozlowski and Bell, 2003; Salas et al., 2008). The increase in multi-authored papers might also reflect the change of journal editorial decisions under a wider collaborative publishing movement. Early studies have already shown that a significant correlation exists between multi-authored submitted papers and the frequency of acceptance for publication (Gordon, 1980). Co-authored papers tended

to have a higher chance of acceptance and higher citation frequencies, so some studies have pointed to the possible relationship between collaboration and quality (Smart and Bayer, 1986).

Although co-authorship growth has been found in all fields, the levels might be different from discipline to discipline. Some disciplines tend to have more multi-authored papers while others might not. For example, it is not unusual for biology articles to have more than 10 authors (Yan and Ding, 2009, p. 21), but this is quite rare for DH articles. In the current dataset, among all the 2,527 collected articles, only 3 articles have 10 or more than 10 authors, i.e., (Baker et al., 2004; Pal et al., 2016; González-Blanco et al., 2017).

Apart from disciplines like biology where multi-authored articles are common, studies have shown that other disciplines also have more extensive level of co-authorship than that of DH. Nyhan and Duke-Williams have pointed out the possibility that the growth of co-authorship in Geography is statistically more significant than that in DH (Nyhan and Duke-Williams, 2014a). Although Geography is considered by some not to belong to the Humanities and Social Sciences (H&SS) or the Computer Sciences where DH is closely related, there are studies examining H&SS fields, demonstrating higher levels of co-authorship than that of DH. For example, Erford et al. investigated the *Journal of Counseling & Development* and found that the average number of authors per article is 2.12 (1994 – 1997) and 2.36 (2006 – 2009) (Erford et al., 2011, p. 78), whereas the current dataset provides 1.60 and 1.91 during the respective periods. This means that DH has fewer number of co-authors compared to the field of Counselling. If Counselling is a field that is not very close to DH, the field of Digital Library might present a stronger argument. During 1994 – 2003, based on data extracted from three Digital Library journals, each paper had a mean of 3.02 authors (Liu et al., 2005), while in the current dataset, during the same period, DH has only 1.83 authors per article.

After comparing with other fields, the current study finds, in line with Nyhan and Duke-Williams, that co-authorship in DH does not seem to be as unique as it was described (Deegan and McCarty, 2012; Nyhan and Duke-Williams, 2014a), e.g., its 'collaborative nature' to distinguish itself from the traditional humanities (Koh, 2012). Because the general academic environment has been encouraging teamwork and embracing more

and more co-authored publications, it is very likely that we can find an increase in co-authorship in all fields, including DH. In general, 'historical periods' is one of the factors that affect the collaborative pattern in DH, although it is not exclusive to DH.

On the other hand, because different historical periods show different levels of co-authorship in DH (i.e., the later the more co-authored papers, as shown in Figure 4.12 and Figure 4.13), can the results clearly indicate that scholars in different generations (not necessarily in the same age) have different level of co-authorship? Studies have shown that older scholars have greater numbers of publications calculating by fractional count (Barjak, 2006; Gonzalez-Brambila and Veloso, 2007; Aksnes et al., 2011), but does this mean that scholars in older generations tend to work alone? The visualised co-authorship networks below present some disagreements (Figure 4.16 and Figure 4.17).

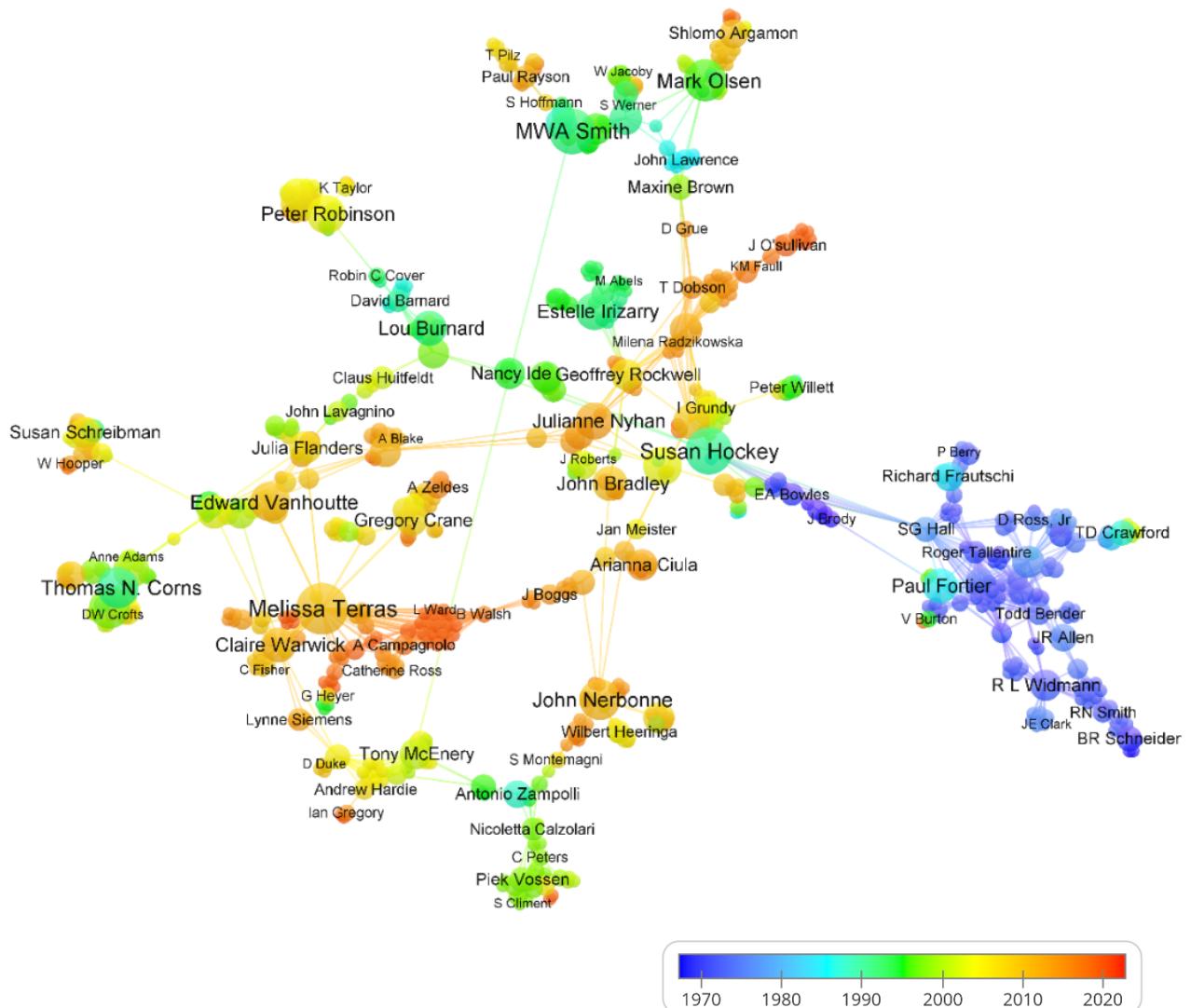


Figure 4.16: DH co-authorship network with average year information, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017), graph created by VOSviewer

Figure 4.16 was presented earlier in this section, and it is referenced here with the same figure number to assist the discussion. It shows the largest co-authorship network in the current dataset consisting of 661 scholars, and the colour of the node indicates the average year of publication. The warmer the colour (i.e., red) the later the year, and the cooler the colour (i.e., blue) the older the average publication year of the author. This network above contains the whole range of the colour palette which means it includes the whole time period of publishing authors (1966 – 2017). Especially on the right of the network, there are a group of authors who published articles mostly in the 1970s, and they are connected to the rest of the network by Susan Hockey who has an average publication year of around 1990. It should be noted that the average number of the publication year does not necessarily imply the particular year that the author was most active. For some authors who are consecutively active in all previous periods, e.g., Susan Hockey (until recently), this number indicates their ‘career plateau’ throughout the time (Chang Boon Lee, 2003, p. 538). Comparatively, some authors are on the rise in this field (in orange and red) while some have been faded out (in blue).

This network visualisation demonstrates that the older generation of scholars also formed apparent co-authorship networks, although its scale is smaller, and its links are limited. The authors who have later and more recent publication years, on the other hand, are spread more widely. In addition, many important bridging nodes (higher betweenness centrality) have publication years later than 1990s or even 2000, e.g., Melissa Terras (2009), Edward Vanhoutte (2009), Julianne Nyhan (2012), John Nerbonne (2010), Nancy Ide (1994).

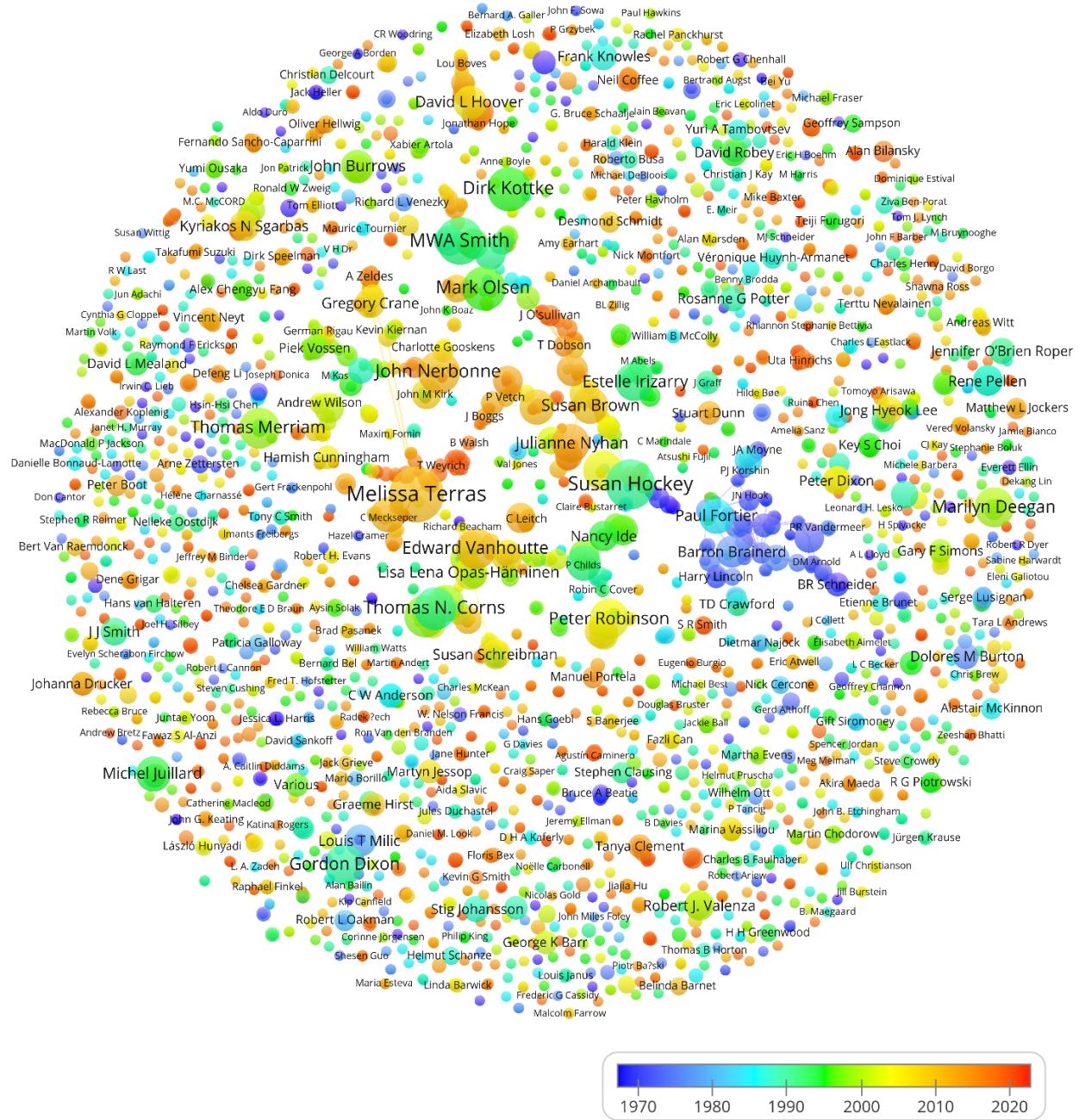


Figure 4.17: DH co-authorship network of all authors with average year information, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017), graph created by VOSviewer

Figure 4.17 above shows the co-authorship network of all 3,382 authors, and as we can see, apart from the largest network of 661 scholars in the middle, the rest have limited links to others and form few noticeable networks. Among all the authors that are not connected to the largest network, all colours of the nodes are spread quite evenly matching the similar time periods' proportions of the total authors in the dataset.

This means that not only are many of the older scholars lone authors, there are many newer ones who also published single-authored papers, scattered around the image disconnected to the main network.

In general, it is too early to suggest that the older scholars publish more single-authored papers, but the longitudinal difference also reflects the importance of using the full counting method to construct a co-authorship network; it narrows the contribution gap between lone authors and team-working authors, and thus eliminates bias when analysing the network structure.

No matter older or newer, in the current dataset, most authors only published single-authored papers (62.01%), and only 19.54% are connected to the main co-authorship network forming a ‘small world’. As reviewed in chapter 2 (Literature Review 2.2.1), many studies have shown that the DH collaborative community might belong to the ‘small world’ type (Nyhan and Duke-Williams, 2014a; De la Cruz et al., 2015; Tang et al., 2017). Although this study has made an effort to construct the largest DH bibliometric dataset, expanding over a longer time period with a greater number, the total of authors connected to the co-authorship network has only grown from 16% by De la Cruz et al., (based on 178 articles) to 19.54% in the current study (based on 2,527 articles). 80.46% of authors were still left outside of the main network for they did not have any co-author links that connected them to the largest cluster. It seems that the expansion of the dataset does not change the ‘small world’ type of co-authorship network in DH, and when compared with co-authorship networks in other disciplines, 19.54% is the lowest. For example, despite that the data samples in these studies vary, the co-authorship network in Medicine contained around 92% of its total authors (Newman, 2001). In Management and Organization, the number was 45% (Acedo et al., 2006). Even Sociology has a higher number of around 34.5% (Moody, 2004).

Combined with the previous studies that have been discussed in chapter 2, we can now conclude that among all the authors who published in DH journals, only a small but strong set of them have been collaborating actively. Within this ‘small world’, everyone knows everyone, but outside of it, this set of authors could only form highly fragmented communication with the rest of the isolated majority. Also, the visualised network has overturned the previous assumption made in this study that under the ‘big

tent' of DH there are 'plural small worlds', as the network only shows one apparent and connected network and the others are small, disconnected and fractured. This indicates that the DH community structure has not grown into complete maturity where most actors are connected (whether loosely or closely).

To help the community become more connected, it is important to further analyse the core connected group of scholars and investigate the factors and influences that could help DH scholars form collaborative relationships.

Although, as this section suggests, 'historical periods' is one of the potential factors that affect the collaborative pattern in DH, it is not exclusive to DH. The following two sections will continue to discuss other factors – gender and country.

4.3.7.2 Gender

Based on the results, there is an apparent gender difference in DH co-authorship networks, and female scholars play an important role in forming the co-authorship connections. In total, there are 2,253 men (66.62%) and 976 women (28.86%), and the gender distribution has been gradually changing over the past 52 years. Figure 4.18 below shows the annual percentages of unique female and male authors, and there is a clear rising trend towards the percentage of female scholars, especially in recent years, although it has still remains the minority.

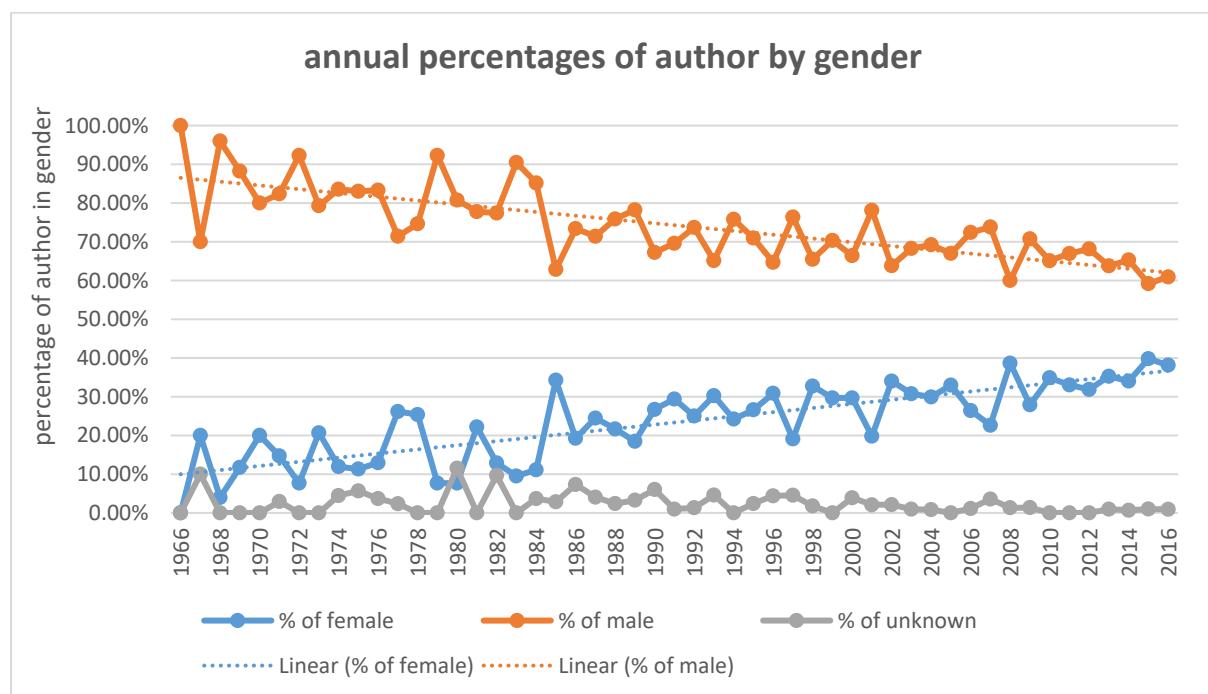


Figure 4.18: The annual percentages of unique number of authors by gender, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017).

Figure 4.18 matches the results of most previous DH gender studies presenting an apparent imbalance between the proportion of female and male scholars (e.g., Weingart, 2012, 2013b, 2014c, 2015e, 2016b; Risam, 2015a). It agrees with previous calls that women are underrepresented not only at the most important gatherings in the field but also in publications (Risam, 2015a; Terras and Nyhan, 2016). However, apart from female scholars being in the minority, the current results also present unanticipated findings about the different collaboration patterns between male and female scholars in DH.

Although the gender distribution in the co-authorship network is similar to that of the total number of authors in the dataset (187 female – 28.29%, 409 male – 61.88%, see below Figure 4.19), it raises interesting questions about the gender differences in forming the DH co-authorship network. Firstly, despite the fact that female scholars account for less than 30%, many of their nodes are functioning as important bridges linking different clusters that would otherwise be disconnected to the main network.

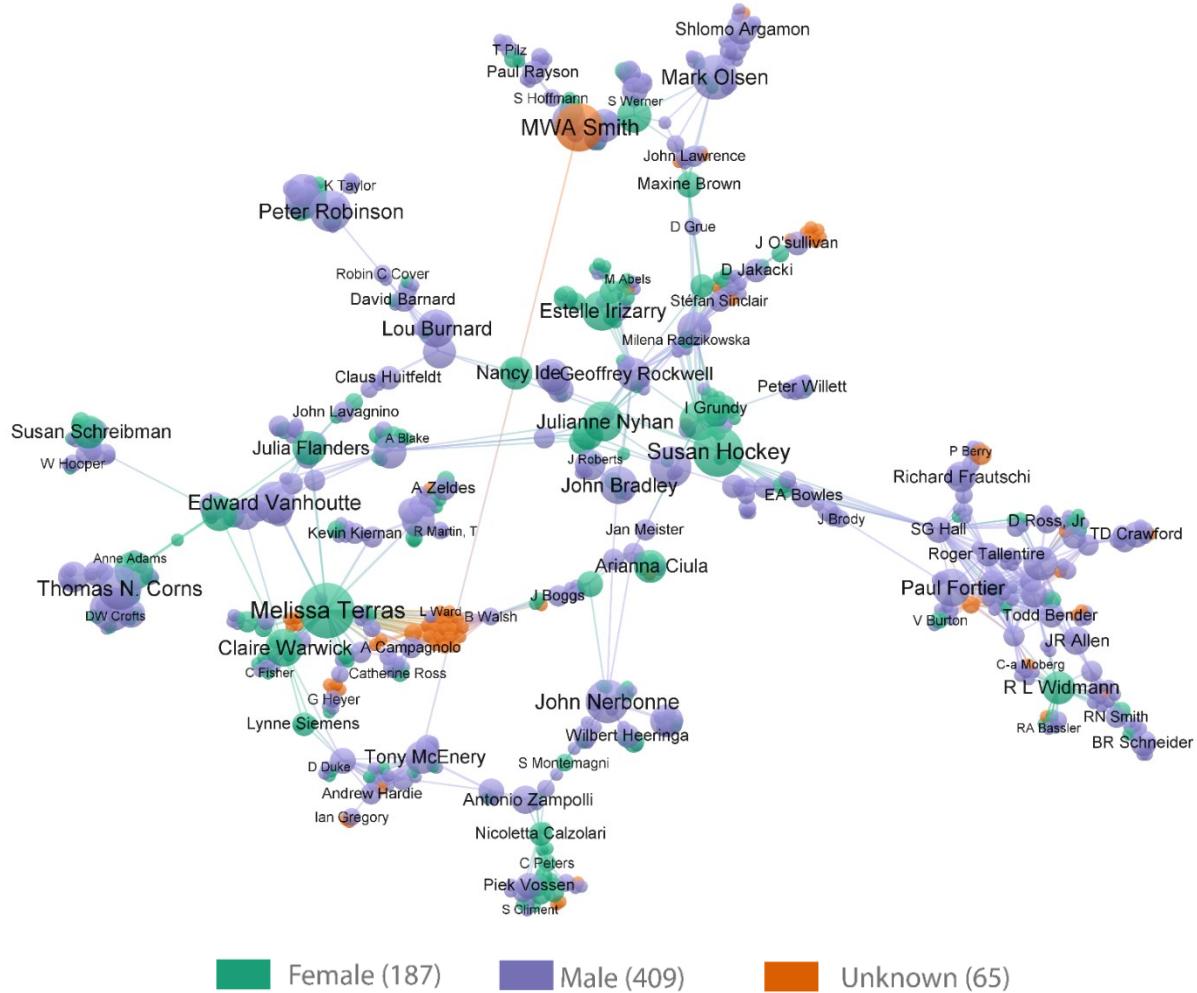


Figure 4.19: The co-authorship network with gender information, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017), graph created by VOSviewer

For example, the node of Susan Hockey links a cluster on the bottom right of the network which would otherwise be disconnected from the main network without its link to her. There are many nodes of female scholars that are playing such important bridging roles, e.g., Susan Brown, Melissa Terras, Julianne Nyhan, Maxine Brown, Claire Warwick, Julia Flanders. Male scholars, on the other hand, are more visibly grouped at the end of each cluster, despite the fact that they take up the majority of places and thus are more visible in general.

In order to justify such bridging roles of female scholars, as mentioned earlier, this study has calculated the *betweenness centrality* of each node (i.e., the number of shortest paths between pairs of nodes) to examine the network structural prominence. It is usually believed that the higher the centrality, the more important the node (Koschützki et al., 2005), and in a co-authorship network, the greater the impact of a

scholar in forming the network (Yan and Ding, 2009, p. 30). Table 4.7 below presents the gender of the top 20 authors ranked by the *betweenness centrality*, and many of them are female scholars (see Appendix B for the complete table).

Table 4.7: The top 20 authors ranked by the highest *betweenness centrality* with gender information (F – female, M – male, U – unknown).

	Author name	Publications	Co-authors	Gender	betweenness centrality (average no. of pairs)
1	Susan Hockey	22	16	F	81501.32
2	Susan Brown	12	25	F	68457.91
3	SG Hall	5	36	U	67345.25
4	Raymond Siemens	8	18	M	57718.00
5	Melissa Terras	27	53	F	49180.26
6	Stan Ruecker	9	31	M	48101.25
7	Julianne Nyhan	13	13	F	36533.59
8	Geoffrey Rockwell	8	23	M	34623.23
9	Maxine Brown	5	9	F	31533.05
10	Paul Spence	3	7	M	30916.79
11	Claire Warwick	11	23	F	30776.83
12	John Nerbonne	16	20	M	27324.34
13	MWA Smith	20	12	U	25812.13
14	Michael Sperberg-McQueen	8	8	M	24844.93
15	Julia Flanders	9	15	F	24471.67
16	Hamish Cunningham	5	19	M	24254.52
17	Mark Olsen	17	21	M	22831.93
18	Tony McEnery	8	21	M	22726.27
19	Whitney Bolton	9	9	F	22691.31
20	Fiona Tweedie	12	14	F	17960.00

As shown in Table 4.7, the value of *betweenness centrality* indicates the average number of pairs that have the shortest paths through the node. Authors have to be connected to at least one person in order to be included. If the value is 0, it means no pair has the shortest path through this node.

According to these results, female scholars are much more important than their male counterparts in forming the co-authorship networks, despite being in the minority. Firstly, there are 9 female scholars and 9 male scholars among the top 20, and the

highest two authors are both female (i.e., Susan Hockey and Susan Brown). Although there are the same number of female and male authors in the top 20 table, one needs to bear in mind that female scholars account for less than 30% of the total number of authors, which means that if they were in the same proportion, there would probably be many more female scholars in the top ranked list. Secondly, among all the 661 authors who are connected to the main network, the average *betweenness centrality* for female scholars is 10,646.99, while it is only 7,542.46 for male scholars. The female average is 41.16% higher than that of the male average, demonstrating that the positions of female scholars in the network are significantly more influential and critical than that of male scholars in DH. Additionally, moving to a broader scale, among all the 3,382 authors in the current dataset, female scholars on average have 2.62 co-authors while male scholars have 2.54. This reflects that women in DH have relatively more collaborators on average than men.

In general, the results show that female scholars have more communications, built more collaborations, and contributed more to the formation of the DH co-authorship network than males. They are not only the main forces to maintain the DH scholarly connections but also the icebreakers who facilitate bridges to isolated groups.

This finding reveals an interesting aspect of the gender difference in the DH community formation. Although we know that female scholars are the minority of the DH community, no previous studies have found the significant contribution they have made to build connections and collaboration from a statistical perspective. On the other hand, this finding is not too surprising when trying to explain such difference and uncover the reasons.

Previous empirical studies have already shown that women in academia actually have more collaborators on average than men (Bozeman and Gaughan, 2011, p. 1393). Female scholars are believed to know more about their networks and to be more knowledgeable in terms of forming their collaboration networks (Brass, 1985). They tend to expand and further develop their individual connections beyond their own areas and specialities, and can thus help to encourage a greater level of interdisciplinary collaboration (Leahy, 2006; Rhoten and Pfirman, 2007). As DH is believed to be more interdisciplinary than many Humanities disciplines, women in DH

may have more opportunities to develop such connections and form more collaborative and diverse networks.

Many studies have revealed that women are better communicators than men and that might be one of the reasons why female scholars in DH have such different collaboration behaviour than male scholars (Peñas and Willett, 2006). In general, women are more ‘expressive’ and ‘tentative’ in conversation, while men are more ‘assertive’ and ‘power-hungry’ (Basow and Rubenfeld, 2003, p. 183). They are believed to be different in their styles of communication, and women often see communication as a way to improve social connections and build relationships while men often use language to express dominance and achieve goals (Leaper, 1991). Thus, many have argued that women tend to be more social in communications while men value their independence (Gilligan, 1993; Chodorow, 1999; Eagly, 2013). Kuhn and Villeva found that significantly more women than men chose teamwork instead of individual tasks, and women were more attracted to collaboration with more optimistic appraisals of their potential teammate’s abilities (Kuhn and Villeva, 2015, p. 115).

Moreover, in the academic environment, female scholars have had many more difficulties than male scholars which might also have helped them to develop their importance in the collaborative network. Numerous studies on female scholars have shown that they have less support, face greater professional isolation and slower rates of promotion, and are more likely to leave academia before gaining tenure (or a permanent post) than their male counterparts (Wasburn, 2007). In order to overcome these challenges, female scholars need to develop exceptional communication skills and make good use of their extensive connections. Collaborations can help female scholars to diminish the lack of ‘social capital’ significantly and better integrate themselves in the academic environment, and co-authorship has influenced their productivity more notably than male scholars (Abramo et al., 2013).

In addition, because female scholars are often the minorities in the scholarly community, they are believed to be more dependent on these networks than their male counterparts who take up more places in academia (Badar et al., 2013). Studies have shown that female scholars are not only benefiting more from their direct and indirect connections within a network, but also benefit significantly from connecting to the disconnected others (Burt, 2005, pp. 30–39). Mehra et al., have pointed out that by

connecting to other isolated nodes that are otherwise disconnected they act like ‘gatekeepers’, and could access more diverse sources and information that can be used to produce more ‘instrumental outcomes’ (Mehra et al., 2001).

However, the motivations for research collaboration and the interpretations of the gendered co-authorship results are various (Ponomariov and Boardman, 2016), and the above analysis offers only a few possible explanations. There are many other discussions to be had. For example, some argued that policies on reducing long entrenched family-related barriers from female scholars might be paying off (Bozeman and Gaughan, 2011, p. 1399). Also, others found that male scholars were more likely to place emphasis on building a solid reputation through independent publications at the beginning of their career, in order to achieve better collaborations later (Abramo et al., 2013, p. 811). Some even disapproved of gender difference and argued that collaboration was less influenced by gender but more by effective leadership and team commitments (Ingram and Parker, 2002).

This section provides a new perspective to understand the gender distribution and difference in DH, and it can help to raise new questions to be examined in future studies. In the following section, this study will continue to uncover other characteristics that the nodes have.

4.3.7.3 Country

Affiliated country distribution could assist in learning the international collaboration patterns in DH. As mentioned, there are 3,382 unique authors, and most of them have been affiliated with English-speaking countries. There are 1,044 US authors (30.87%), 461 UK authors (13.63%), 199 Canadian authors (5.88%), 64 Australian authors (1.89%). There are also many authors who have affiliation in Germany (187 individuals, 5.53%), France (151, 4.46%), the Netherlands (117, 3.46%), Italy (103, 3.05%), Spain (74, 2.19%), and Japan (72, 2.13%). Figure 4.20 below shows the number of unique authors in each affiliated country (top 15), and the complete table of all the 62 countries can be found in Appendix B.

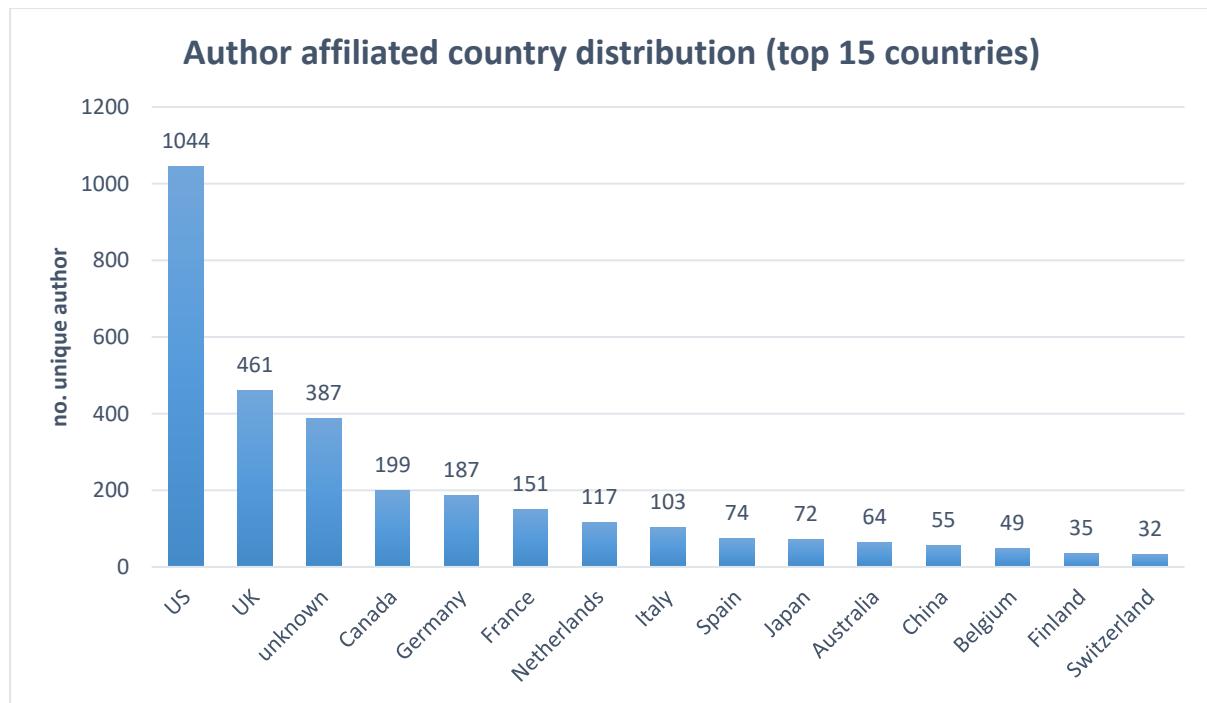


Figure 4.20: The number of unique authors in each affiliated country (top 15), data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017).

Figure 4.20 provides a general affiliated country distribution based on publications in *CHum*, *DSH/LLC* and *DHQ*, and it agrees with the country distribution conducted by previous studies, e.g., ADHO annual attendance (Weingart, 2012, 2013b, 2014c, 2015e, 2016b), as well as DH journal publications (De la Cruz et al., 2015; Tang et al., 2017). As many studies pointed out that DH is a predominantly Anglophone environment (Galina, 2014; Fiornonte, 2017; Mahony, 2018), movements such as diversity, inclusivity, decolonisation, intersectionality, and openness have been called for (Gil and Ortega, 2016; Weingart, 2016a; Risam, 2018). While the communities have been trying to improve the representation and diversity, it is important to turn the focus back on the current dataset to study the forces that encourage such movements and international collaboration.

From the perspective of the number of articles, as mentioned, among the 2,527 articles in the current dataset, 960 of them are multi-authored (37.99%) and 244 articles have been co-authored internationally (i.e., articles published with scholars from more than one country, 9.66% of the total articles), and this figure does not count the cases when an individual scholar affiliates with multiple countries. Figure 4.21 below shows the

annual share of the international co-authored articles in line graph (please see Appendix C for more information).

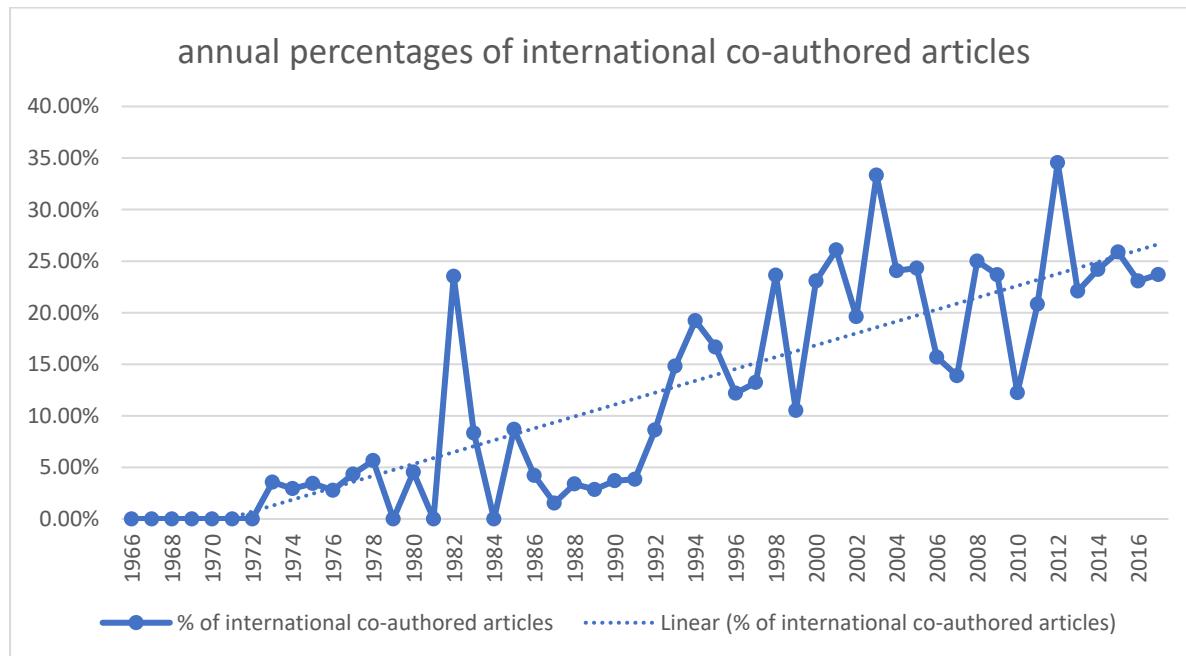


Figure 4.21: The annual percentages of the international co-authored articles, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017).

As shown in Figure 4.21, although fluctuating, the proportion of international co-authored articles has been growing generally during the 52-year period, and there is an obvious rising trend for more international collaborations in DH. It was 0% from 1966 to 1972, gradually increased to 8.7% (1985), 12.2% (1996), 23.08% (2000), 25% (2008), peaked at 34.55% (2012), and finally reached 23.7% (2017). Although the average mean remains at 9.66%, the numbers in the later period (especially after 2000) are higher than numbers in other disciplines during the same periods. For example, based on a comprehensive range of globally published Science and Engineering journals in Scopus, the international co-authored papers only accounted for 15% in 2000 (US National Science Board, 2000), while in DH, it was 23.08%. The figure rose to 17% in 2008 based on the US report (US National Science Board, 2020, p. 15), while it was 25% in DH. Although there is no DH data for 2018, the DH proportion in 2017 was 23.7%, which is still slightly higher than 23% – the US figure in 2018 (US National Science Board, 2020, p. 15). Considering that co-authorship in Science and Engineering fields should normally be more frequent and international (Gläzel and

Schubert, 2005a), the continuous higher proportion of internationally co-authored articles in DH presents a surprisingly worldwide and diverse collaboration pattern.

Moreover, the DH international co-authored share is notably higher than many fields in Social Sciences, too. Henriksen has analysed the international shares of co-authorship based on 4.5 million articles published in 1980–2013 indexed in the Social Science Citation Index (SSCI), Web of Science (Henriksen, 2016). Figure 4.22 below is the comparative line graph combined with the results of DH by the current study and the percentages published in Henriksen's study.

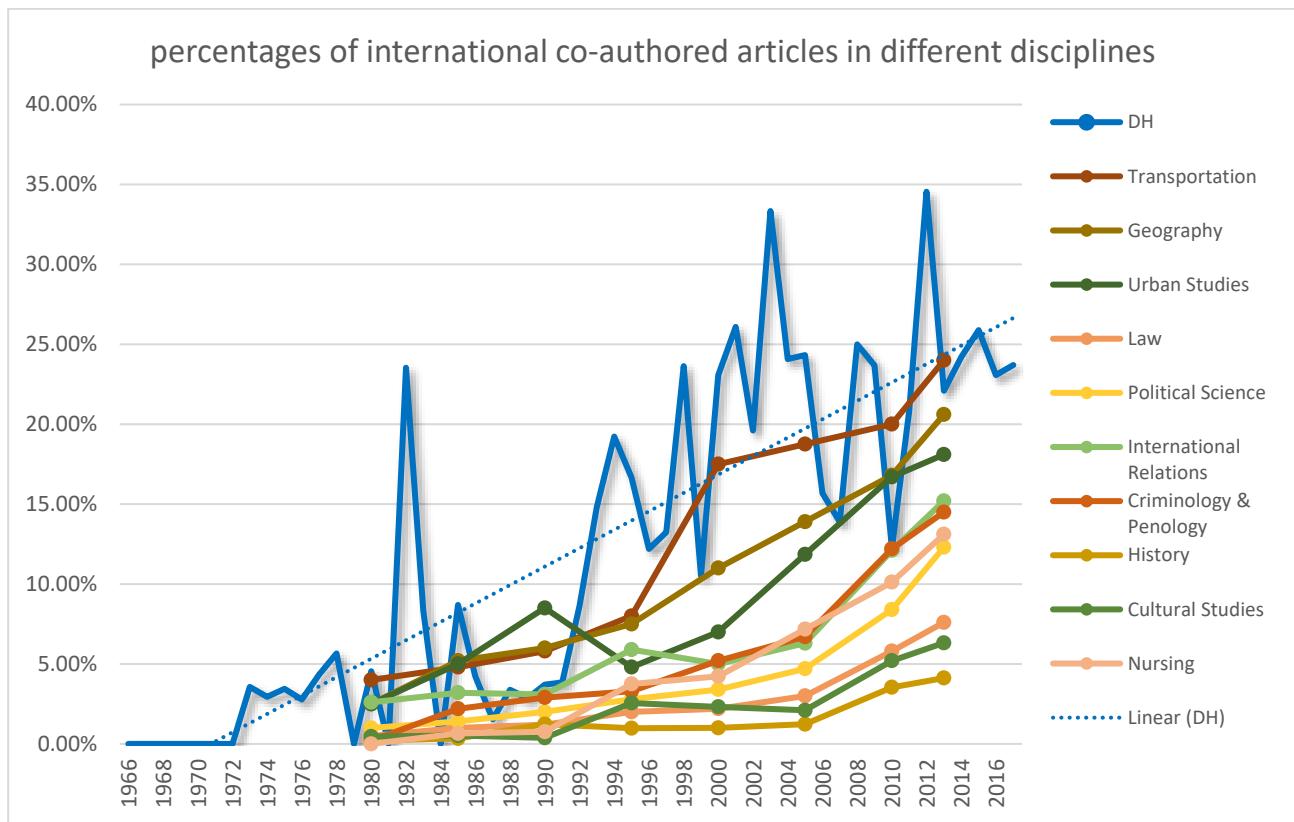


Figure 4.22: The annual percentages of the international co-authored articles in DH combined with the percentages of other disciplines 1980 – 2013 in Henriksen's study (Henriksen, 2016).

As shown in Figure 4.22, although the DH percentage fluctuates over the period due to the different scales of the two studies, its percentage values are generally higher than most of the disciplines during Henriksen's study (1980 to 2013), such as Transportation, Geography, Urban Studies, Law, Political Science, International Relations, Criminology & Penology; and fields like History, Cultural Studies and

Nursing are among the lowest (Henriksen, 2016). The figure indicates a much greater expansion of DH research collaboration than other fields around the world. Although the overall international share of co-authorship is shown to be rising in all disciplines (Glanzel, 2001), DH is one of the fields with considerably more global collaboration. While DH scholars may not collaborate as frequently as those in other disciplines, when they do so the collaborations tend to be more international than in other disciplines.

Some have claimed that international co-authorship often produces papers with higher citation rates and possibly higher impact too (Glanzel, 2001; Glanzel and Schubert, 2001; Sugimoto et al., 2017), even though national collaborations across industries (i.e., academic, government and industry collaboration) also have higher citation and impact (Perkmann and Walsh, 2009; Frenken et al., 2010). To investigate how to improve the level of international co-authorship in DH, it is important to research what types of DH authors have contributed the most. Figure 4.23 below demonstrates the different types of authorship in each affiliated country. In particular, the green areas indicate the number of scholars who have internationally co-authored (number label as shown in chart).

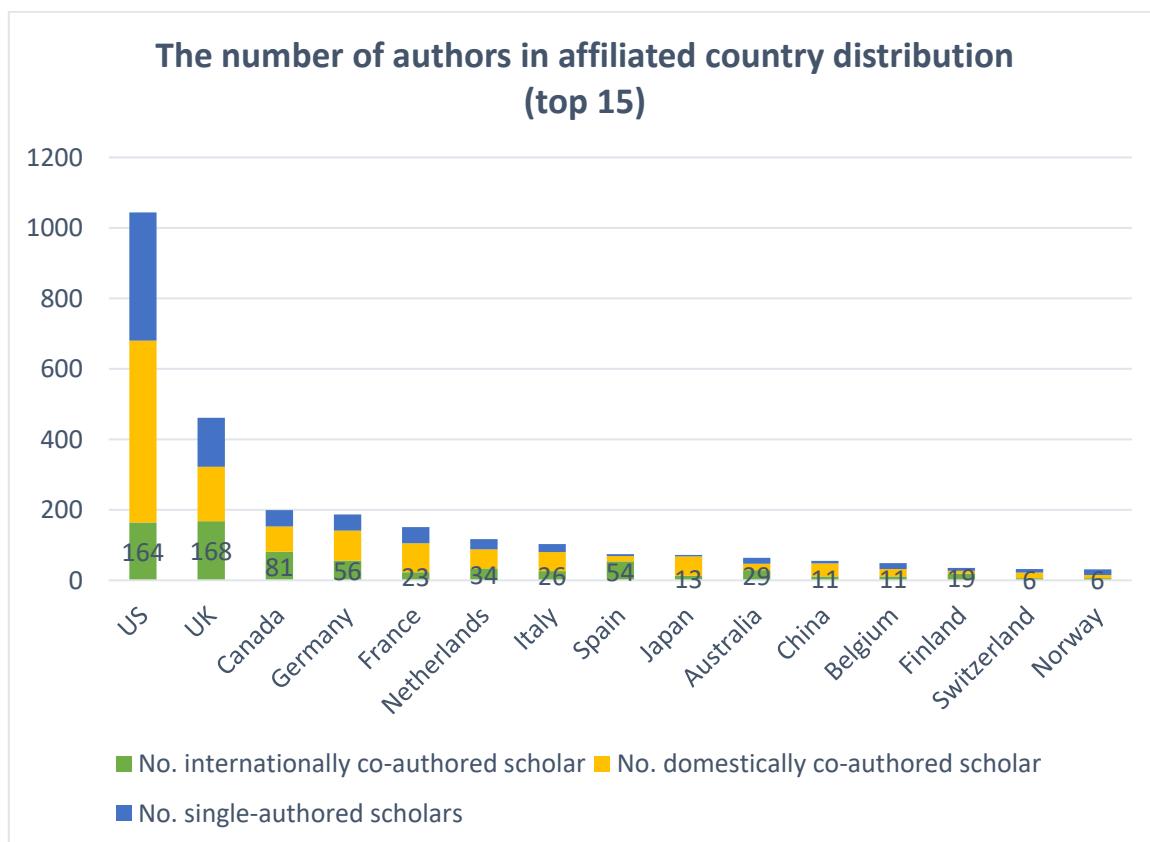


Figure 4.23: The different author distributions in each affiliated country (top 15), data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017).

As shown in Figure 4.23, countries, such as the UK, the US (despite low in proportion), Canada, Germany, and Spain (significantly high in proportion) contribute the most internationally co-authored articles. Although the US, despite having the highest number of scholars and co-authored scholars in this dataset, only ranks second, 164 (after the UK, 168), when it comes to the number of international co-authoring scholars.

Table 4.8 shows the top 15 countries with the most scholars in the dataset, and they are ranked by their international co-authored rate. Columns from left to right present the number of unique scholars, the number of co-authored scholars (i.e., scholars who co-authored articles), the number of international co-authored scholars (i.e., scholars who have co-authored an article with collaborators affiliated in at least two countries, excluding individuals who have multiple affiliations), the percentage of co-authored scholars (i.e., number of unique scholars divided by the number of co-author scholars), and the percentage of international co-authored scholars (i.e., number of unique scholars divided by number of international co-authored scholars) in the current dataset. For the complete table, please see Appendix C.

Table 4.8: The top 15 countries with the most scholars ranked by the international co-authored rate, data extracted from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

		No. unique scholar	No. co-authored scholar	No. international co-authored scholar	Co-authored %	International co-authored %
1	Spain	74	69	54	93.24%	72.97%
2	Finland	35	27	19	77.14%	54.29%
3	Australia	64	47	29	73.44%	45.31%
4	Canada	199	153	81	76.88%	40.70%
5	UK	461	322	168	69.85%	36.44%
6	Germany	187	141	56	75.40%	29.95%
7	Netherlands	117	88	34	75.21%	29.06%
8	Italy	103	80	26	77.67%	25.24%
9	Belgium	49	32	11	65.31%	22.45%
10	China	55	48	11	87.27%	20.00%
11	Norway	31	15	6	48.39%	19.35%
12	Switzerland	32	22	6	68.75%	18.75%
13	Japan	72	68	13	94.44%	18.06%
14	US	1,044	680	164	65.13%	15.71%
15	France	151	105	23	69.54%	15.23%

From Table 4.8, although the sample size of each country varies, we can see that 72.97% of scholars affiliated in Spain have co-authored articles with scholars from other countries making Spain the most internationally collaborative country in the current dataset. Finland, Australia, Canada and UK are ranked the second to the fifth, respectively (54.29%, 45.31%, 40.70%, 36.44%). However, there are some interesting cases with high co-author rate but low international co-author rate, i.e., frequent co-author activities but often limited to domestic scale. For example, Japan has the highest co-author rate (meaning that 94.44% of Japanese scholars have co-authored articles), but its international co-author rate is the third from the last among the 15 countries (suggesting that only 18.06% of Japanese scholars have ever published with non-Japanese scholars). Although the US has the highest number of scholars (30.87% of the total) and the second highest number of scholars who have co-authored internationally (i.e., 164), and its co-authored rate is not particularly low (65.13%), only 15.71% of the Americans have ever co-authored articles with scholars outside of the US and rank in the penultimate. Similarly, scholars affiliated in France, Switzerland,

and China also have relatively high co-author rates but low international co-author rates.

Although, surprising to some, such country ranking is similar to the previous one provided by the US National Science Board in 2020. Based on a wide range of Science and Engineering journals in Scopus, they showed that the most internationally collaborative countries are the UK (62%), Australia (60%), Canada (56%), Germany (53%) and Spain (53%), and similarly, the US has an international co-authorship percentage of 39% which is below the average (US National Science Board, 2020, p. 15).

Why do the majority of American or Japanese (or French, Chinese, etc.,) scholars choose to actively collaborate but just with scholars of the same country? At the same time, why do people affiliated in e.g., Spain, Canada, the UK, Germany and the Netherlands, choose to collaborate more frequently with scholars in other countries? There are many aspects to unpack these questions, e.g., funding bodies, international relations, languages, policies, personal connections, etc., which future studies could address. The current study focuses on finding which country they like to collaborate with based on co-authorship, and it seems that language plays a key part.

Table 4.9 below shows the most frequently co-authored country pairings. It is not surprising to find that English-speaking countries, especially US and UK, form most of the co-authorship country pairs, because the article data was collected from English-language journals. However, different Anglophone countries seem to collaborate differently. For example, scholars in the US often form authorship partnerships in English-speaking countries (e.g., the UK, Canada, Australia), while those in the UK not only collaborate with scholars in English-speaking countries, but also in non-Anglophone countries more often than the US (e.g., Germany, the Netherlands, China, Finland, Japan).

Table 4.9: The top 20 most frequently co-authored international country pairs, data extracted from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

Rank	Country pair	No. co-authored articles	Rank	Country pair	No. co-authored articles
1	UK - US	36	11	Netherlands - Germany	7
2	US - Canada	29	12	UK - China	6
3	UK - Canada	19	13	Netherlands - Belgium	5
4	UK - Germany	19	14	Netherlands - Italy	5
5	US - Australia	12	15	Germany - Canada	4
6	US - Germany	10	16	UK - Finland	4
7	US - Ireland	10	17	UK - Ireland	4
8	UK - Australia	8	18	Spain - Italy	4
9	UK - Netherlands	8	19	UK - Japan	4
10	US - Netherlands	8	20	Spain - Netherlands	4

Moreover, in terms of domestic co-authorship patterns, scholars who are affiliated with a UK institution tend to collaborate more often than those of other countries. Table 4.10 below shows the top 20 countries with domestically co-authored publications in this dataset, and one where the UK has more domestically co-authored articles than the US. One also needs to bear in mind that the number of US authors in this dataset (1,044 individuals) is more than double that of UK authors (461).

Table 4.10: The top 20 domestically co-authored countries, data extracted from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

Rank	Country pair	No. co-authored articles	Rank	Country pair	No. co-authored articles
1	UK - UK	107	11	Italy - Italy	12
2	US - US	106	12	China - China	10
3	Canada - Canada	57	13	South Korea - South Korea	8
4	Germany - Germany	47	14	Finland - Finland	6
5	Netherlands - Netherlands	32	15	Norway - Norway	6
6	Australia - Australia	21	16	Japan - Japan	5
7	Spain - Spain	18	17	Singapore - Singapore	5
8	Belgium - Belgium	13	18	Switzerland - Switzerland	5
9	France - France	12	19	Israel - Israel	4
10	Ireland - Ireland	12	20	Sweden - Sweden	4

If we see the collaboration from the perspective of network analysis, the measure of *betweenness centrality* offers a clearer interpretation. Figure 4.24 shows the co-authorship network colour-coded with affiliated country information, and Table 4.11 provides the top 15 countries with the most scholars ranked by the *betweenness centrality*.

The affiliated country distribution on the co-authorship network is somewhat similar to that of the total number of authors in the dataset, but we can see the clear increase in the proportion of UK authors. US authors account for 31.16% of the nodes on the network, and this figure is similar to the proportion of US authors in the total number of collected authors (30.87%), while there are 18.91% of UK authors on the network, and this number is higher than its percentage in the general author collection (13.63%). This difference indicates that there is a higher proportion of UK authors connected to the co-authorship network. Figure 4.24 below shows the co-authorship network colour-coded with affiliated country information.

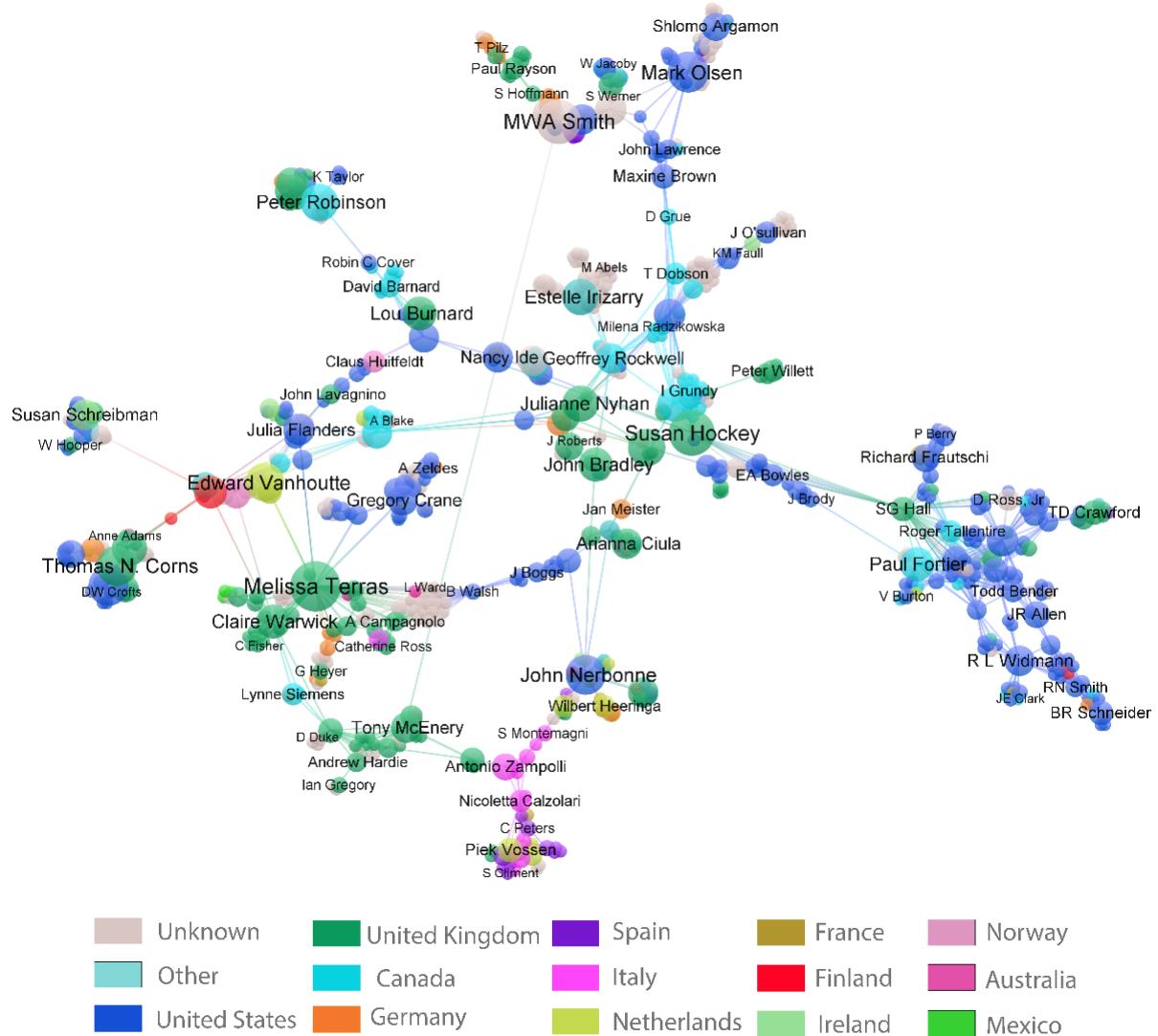


Figure 4.24: The co-authorship network with affiliated country information, data from journals *CHum*, *DSH/LLC* and *DHQ* (1966 – 2017), graph created by VOSviewer

Table 4.11: The top 15 countries with the most scholars ranked by the betweenness centrality, data extracted from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2017.

		total scholar	co-authored %	international co-authored %	no. scholar on the network	betweenness centrality (average no. of pairs)
1	Belgium	49	65.31%	22.45%	2	7,521.57
2	Canada	199	76.88%	40.70%	54	4,233.65
3	UK	461	69.85%	36.44%	126	3,713.76
4	Italy	103	77.67%	25.24%	14	2,715.79
5	US	1,044	65.13%	15.71%	206	2,088.80
6	Norway	31	48.39%	19.35%	4	1,884.98
7	Germany	187	75.40%	29.95%	24	1,036.80
8	Spain	74	93.24%	72.97%	16	926.55
9	Netherlands	117	75.21%	29.06%	15	733.74
10	France	151	69.54%	15.23%	8	0
11	Japan	72	94.44%	18.06%	1	0
12	Australia	64	73.44%	45.31%	4	0
13	China	55	87.27%	20.00%	1	0
14	Finland	35	77.14%	54.29%	7	0
15	Switzerland	32	68.75%	18.75%	0	0

Although publishing in the three selected DH journals is each scholar's individual choice, such choice is greatly influenced by various factors, and national wealth is believed to be a significant one (Ammon, 2006, p. 7). Despite that some have questioned the idea (Trajtenberg, 1990; Harvey and Green, 1993), many have referred to De Solla Price's argument about the relationship between the number of academic publications and national wealth (Stephan, 1996; Nickerson, 1998; Börner et al., 2005), which is:

[...] the share each country has of the world's scientific literature by this reckoning turns out to be very close—almost always within a factor of 2—to that country's share of the world's wealth (measured most conveniently in terms of GNP). The share is very different from the share of the world's population and is related significantly more closely to the share of wealth than to the nation's expenditure on higher education. (De Solla Price, 1986, p. 142).

Although at this stage it is difficult to justify this claim in the DH environment, we can see that countries having the most DH publications are among the wealthiest, e.g., the US, UK, Canada, Germany, France, the Netherlands, Italy, Spain, Japan (see Figure

4.20 and Figure 4.24). In particular, as the economically strongest language community in the world (Ammon, 2006, p. 7), Anglophone countries as a whole have the most DH publications, as well as co-authored and internationally co-authored publications, which might be due to the fact that the selected journals are mainly published in English. However, compared to the UK (36.44%) and Canada (40.70%), why does the US have the second lowest internationally collaborative rate (15.71%)?

Some believed that it is a problem of scientific size and suggest that the more scholars a country has, the less need there is for international collaboration, as Melin suggested:

It is generally assumed that there is a negative correlation between national scientific size and amount of international research collaboration: The larger the size is of the national scientific arena, the lesser the amount of international research collaboration. (Melin, 1999, p. 161)

In 2019, the UK has a population of 66.44 million, while Canada has around 37.59 million people; the US, on the other hand, has almost ten times the population of Canada (327.2 million in 2018) (Roser et al., 2020). 42% of the UK population have higher education qualifications (UK Gov, 2017), and 45.16% in the US (US Census Bureau, 2018, p. 2018). Fewer scholars in the UK may lead to a lower rate of domestic collaboration as they do not have as many options as US scholars when selecting a domestic collaborator (Glänsel and Schubert, 2005b), and it seems to be the same with Canada (US National Science Board, 2010). Thus, it is understandable that more than half of UK research is produced through international collaborations (Universities UK, 2018), while the US is often among the countries with the lowest international collaboration rate (US National Science Board, 2000, 2010, 2020).

Nevertheless, population differences do not always have the ability to explain different patterns of international collaboration. The cases of high collaboration rate but low international rate in DH is not only associated with the US; Japan has the largest gap (94.44% co-author rate but only 18.06% international rate), followed by China (87.28% co-author, 20.00% international) and France (69.54% co-author, 15.23% international) (as shown in Table 4.8). Is it also because these countries have a large population size? Although China and Japan are known for their high population density, their low international co-authorship rates might depend more on their languages and geographic locations rather than population size. Because of the difficulties of learning

a foreign language, Chinese scholars may be more comfortable working with Chinese-speaking scholars, and many international collaborations in China are found to be with Chinese immigrant scholars in foreign countries although published in English (Wang et al., 2013). The international collaboration in Japan, too, relies heavily on its geographic location and language, and Japanese scholars are noticed to collaborate more often with East Asian scholars (whose language and location is closer) than with Westerners (Miquel and Okubo, 1994, p. 286; Zitt et al., 2000, p. 639).

As for France, it appears to be impossible to use population size to explain their low international co-authorship rate for they do not have a relatively large population nor high population density. While the combination of geographic and (partial) linguistic proximity can help to interpret the existing collaborations between French-speaking countries (e.g., France, Switzerland, Belgium, and Canada), why does France in particular have the lowest international rate among the top 15 productive countries in DH? Despite the French stereotype of not speaking English, although some claimed otherwise (Eurobarometer, 2012, p. 37), many scholars have discussed its low collaboration rate with other countries such as the UK, the US, Germany, Japan, and China (Okubo et al., 1992; Zitt et al., 2000, p. 636; He, 2009). As Ammon showed, French is a type of international language that has suffered the most from the Anglophone globalisation when comparing their present with their previous situation (Ammon, 2006, p. 16). Unlike speakers of languages that have never become international, Francophone scholars have arguably not yet fully adjusted to the new Anglophone dominance, especially the older generation ‘who have suffered a dramatic social decline’ (Ammon, 2006, p. 16).

In contrast to French, German had been ‘boycotted’ long before English became the international language that it is now (Bailey et al., 1986; Ammon, 2006, p. 7; Ferguson et al., 2011). The ‘systematic exclusion of the German language from international conferences and publications’ was partly due to the historical and political motivations after World War I (Schroeder-Gudehus, 1990; Ammon, 2006, p. 7). German scholars were forced to use other languages to publish and encouraged the use of English (Reinbothe, 2013). Therefore, it is not surprising to find that Germany is among the countries with high international rates (29.95%). This is similar to the situation where

countries with non-international languages try to reach an international audience (e.g., Finland, the Netherlands, Italy, Belgium). As Ammon explained:

They have always had to communicate internationally in a foreign language. For them the present situation even has the advantage that they are no longer forced to acquire skills in several languages of science as was the case formerly, but can, with perhaps slight restrictions here and there, limit their endeavours to a single language. They therefore are not deeply worried about the hegemony of English as the international language of science but even consider it an advantage. (Ammon, 2012, p. 16)

Therefore, it is not difficult to see why Finland (54.29%), the Netherlands (29.06%), Italy (25.24%), and Belgium (22.45%) have relatively higher international collaboration rates in DH, too.

However, international collaboration is a complicated phenomenon that depends on a variety of complex factors, such as the direct benefits (e.g., language convenience, equipment and material advantages) and indirect benefits (e.g., strategic, economic, political, or other formal policies) (Georghiou, 1998). Population, language and country are but only three indicators and cannot explain the DH international collaboration pattern thoroughly. This study is an initial investigation on the DH international collaboration which could be further expanded in the future.

On the other hand, because the international co-author rate is influenced by various factors that often cannot be adjusted by individual scholars, there is no right or wrong to having a high or low international co-author rate (Glänzel and Schubert, 2005b, p. 336). Although there is an apparent (but limited) correlation between citations and international collaboration, it is not a direct measure of research quality (Schmoch and Schubert, 2008). Moreover, many DH subjects and topics are language-specific, and it is naturally difficult to conduct such collaboration among international scholars who speak different languages. In addition, the samples of each country in this study varies. For example, in Table 4.11, Belgium has the highest *betweenness centrality* because it only has two scholars in the current sample, and one of them is the editor-in-chief of *DSH/LLC*, Edward Vanhoutte, who coordinates publications and potentially foresters collaborations even if unintentionally. This is why Belgium has the highest central position on the network. On the contrary, countries such as the UK and the US have

126 and 206 scholars on the network, respectively, and the number of US scholars is 100 times the number of Belgian scholars. Despite the fact that many US and UK scholars have higher *betweenness centrality* than Vanhoutte, their country average values are lower than Belgium. Future studies could work on expanding the dataset and normalising the analysis.

5 DH Twitter Network Analysis

Although the *subject specialty* presents visual networks of the DH intellectual structure based on publications, it does not cover every aspect of the multifaceted DH invisible college. Studies of *social actors*, on the other hand, can provide a different overview and reveal scholarly networks that are unavailable to bibliometric analysis. With the technology development on social media, sociometric studies provide a ‘material mirror’ that can be used to decipher the complexities of online communities (Burton, 2015, p. 5).

Compared to the formal communication channel, a large part of the DH scholarly activities can be found by tracing informal communications where scholars build connections, establish collaborations and exchange ideas outside of publications. Before the formal research link (e.g. co-authorship, citing and cited connections) was set up by any published work, many informal activities might have led up to and formed possibilities for such links: for example, how do co-authors know each other; how do they communicate their research interests; how do they build collaborative projects; how do they co-publish their academic works; and how do others read and cite these works? As discussed in chapter 3 Methodology, with the help of Twitter analysis, one can learn more about the DH scholarly connections via social media.

From the social analysis perspective, therefore, this chapter formulates the original research questions within the sociometric context as the following:

- a) *Subject*: What are the main DH topics on Twitter? How do they relate to each other? How do they develop over time?
- b) *Scholar*: Who are the influential DH scholars on Twitter? How interactive are they? What patterns of social interactions can be identified based on their retweet activities? What might be the determining factors of the structure?
- c) *Environment*: How diverse are the backgrounds of DH scholars (i.e. gender, affiliated country) on Twitter? How do gender and country diversities influence the two questions above (i.e. the Twitter topic and community structure)?

This chapter provides a series of detailed steps on what was done to construct two Twitter networks to explore the DH *subject*, *scholar*, and *environment* – the hashtag co-occurrence network and the user co-retweet network. In general, the former was constructed based on the co-occurrence of hashtags extracted from individual tweets while the latter was constructed by the co-retweet activities of the DH Twitter users.

Firstly, this chapter introduces the compilation of the DH Twitter dataset that was used for both networks (5.1). User profiles as well as their publicly posted tweets were collected from Twitter (2006-2017). Then, the network construction procedures of the two social networks are demonstrated separately in section (5.2) and (5.2.6) to give thorough descriptions and explanations of the research operation. Both studies had similar steps to the bibliometric networks in chapter 4, but each had a distinct research aim and emphasis. Lastly, a brief summary of how they could help to answer the research questions (section 5.3.6) is demonstrated at the end.

It needs to be noted here that because of the nature of the metadata collected by using the Twitter API, in general, a ‘tweet’, as a noun in this study, indicates any original tweet, retweet, quote (or comment) retweet, and reply, and unless specified. The term ‘tweet’, as a verb, refers to any of the activities of ‘posting a tweet’ (e.g., tweet, retweet, quote retweet, reply, etc.,); these distinct kinds of tweets and activities are differentiated and explained when calculating them separately in later sections.

Similar to bibliometric networks, when discussing Twitter networks, terms such as ‘node’, ‘actor’, ‘user’, and ‘scholar’ have been used interchangeably, and terms like ‘edge’, ‘link’, ‘relationship’, ‘co-occurrence’ (for co-hashtag study), and ‘co-retweet’ (for co-retweet study) have also been used interchangeably.

5.1 Data Collection and Cleaning

As discussed in the earlier chapters, the DH community was believed to be highly active on Twitter, but in what way could one identify the DH community (or communities) there? How does one select the representative DH users? Who should be included and who should not?

Different methods have been applied by previous Twitter user studies to identify representative users. Some selected users by analysing biography descriptions (or

user bio) where the user provided a brief narrative about themselves and/or their interests (Grandjean, 2016, p. 2; Song et al., 2016, p. 8). Normally, this method would be conducted by searching the keywords that are related to the subject within the Twitter bios; for example, searching ‘digital humanities’ or ‘humanities computing’ to retrieve a list of DH-related Twitter users who include any of these terms in their bios. Although with methods to enhance the selection procedure and accuracy, this method is mainly based on the texts of user bios so it might fail to pick up the active and influential users who have short and/or irrelevant descriptions, those who use languages other than English (or the language of the keywords) on their bio statements, or be unable to exclude irrelevant users from other professions who happen to match the search requirements.

For example, at the time of data collection (5th November 2017), scholars like Susan Schreibman (@schreib100)³⁹ and Gregory Crane (@PhilologistGRC)⁴⁰ did not have any content on their Twitter bio descriptions. Jim Groom (@jimgroom)⁴¹ had ‘a b twit’ as the bio, while Chuck Rybak (@ChuckRybak)⁴² had ‘No Dream Deferred’, and Barbara Bordalejo (@bordalejo)⁴³ had ‘She edits, writes, reads and has opinions’. Shigeki Moro⁴⁴ had ‘花園大学教授 / 『論理と歴史－東アジア仏教論理学の形成と展開』（ナカニシヤ出版、2015）<http://www.nakanishiya.co.jp/book/b196289.html> …など。⁴⁵ in Japanese as his bio description. These scholars are active DH knowledge contributors but applying such data collection methods will miss them.

³⁹ Susan Schreibman, Professor of Digital Humanities and Director of *An Foras Feasa*. <https://www.maynoothuniversity.ie/people/susan-schreibman>

⁴⁰ Gregory Ralph Crane, Alexander von Humboldt Professor of Digital Humanities. <http://www.dh.uni-leipzig.de/wo/gregory-crane/>

⁴¹ Jim Groom, the co-founder of Reclaim Hosting, Previously was the director of the Division of Teaching and Learning Technologies and adjunct professor at the University of Mary Washington. <http://jimgroom.net/about/>

⁴² Chuck Rybak, professor and interim Dean at University of Wisconsin-Green Bay. <https://www.uwgb.edu/english/faculty-staff/rybakc/>

⁴³ Barbara Bordalejo, a textual critic, editor, digital humanist, and an elected member of the Executive Committee of EADH. <http://eadh.org/barbara-bordalejo>

⁴⁴ Shigeki Moro, professor at Faculty of Letters, Hanazono University. <https://researchmap.jp/moroshigeki/>

⁴⁵ Japanese to English translation: Professor / Hanazono University / Logic and History - Formation and Development of East Asian Buddhist Logic’ (Nakanishiya Publishing, 2015) <http://www.nakanishiya.co.jp/book/b196289.html> ... etc etc.

Some studies realised this problem and used lists of researchers compiled and maintained by other people to collect Twitter data instead.⁴⁶ For example, Dan Cohen set up a ‘comprehensive list of scholars in digital humanities & editors of Digital Humanities Now @dhnow’, and, at the time of writing, it included 346 users (Cohen, 2009, 2019). Marin Dacos created a list of 37 DH Twitter users (Dacos, 2019), and the University of California, Berkeley maintain a list of ‘open community of people from UC Berkeley involved in digital humanities (broadly conceived)’ on Twitter (DH at Berkeley, 2019). The data collection based on such lists, however, also has the potential to introduce bias in how users are added to the lists because they are maintained based on an individual’s knowledge and inspection.

Nevertheless, to find a definitive list of the DH community who are on Twitter is practically impossible, and there are no clear boundaries between Twitter users inside and outside of DH. Moreover, similar to bibliometric data collection, the question of finding a DH user list in itself brings this study back to its own research questions about ‘who we are’.

This study, therefore, firstly extracted the user list based on the two bibliometric networks created in chapter 4. The idea was to extend the two author lists (cited author and published author) that were generated from the co-citation and co-authorship networks to match the DH Twitter community. An experimental data collection was conducted in June 2017 (from 2017-06-02 to 2017-06-30). The collection task was initially focused on manually finding Twitter accounts for the most cited 527 scholars (where *fractional nonself citation* value was equal to or greater than 7.0) from the first list (co-citation). However, 63.76% (336) of them were deceased and only 9.87% (52) of the cited scholars had identifiable Twitter accounts. This proportion of cited authors was not enough for constructing a representative social network. From the second list (co-authorship), the Twitter information of the 665 most productive authors who published the most articles based on the co-authorship data (where the number of publications was equal to or greater than 2) was checked manually. Still, it turned out that the collected data was not suitable for further Twitter analysis. There were only

⁴⁶ For instance, Côté and Darling used an online list of ecology and evolutionary biology (EEMB) researchers ‘curated’ by Byrnes (Côté and Darling, 2018), and collected 450 users to represent ecologists and biologists on Twitter. Such Twitter user lists also exist for DH.

44.06% (293) of the authors that could be identified on Twitter but 38.57% (113) of these Twitter users were not active and had tweeted fewer than 10 times in total. The low overlap between authors on publication lists and scholars on social media is also found in other empirical studies; Holmberg and Thelwall had the same problem when they collected DH twitter accounts from highly cited DH scholars (Holmberg and Thelwall, 2014).

To practically include more representative DH users, therefore, this study identified users from the following list of seven influential DH organisational accounts, i.e., ADHO and its member organisations (see Table 5.1).

Table 5.1: The selected organisation accounts and the number of users they followed at the time of data collection.

Acronym	Organisation full name	Twitter handle	No. users it followed
ADHO	The Alliance of Digital Humanities Organizations	@ADHOrg	977
EADH	The European Association for Digital Humanities	@eadh_org	506
ACH	Association for Computers and the Humanities	@achdotorg	889
CSDH/SCHN	Canadian Society for Digital Humanities / Société canadienne des humanités numériques	@csdhschn	409
centerNet	centerNet	@DHcenterNet	1,023
aaDH	Australasian Association for Digital Humanities	@aaDHumanities	400
JADH	Japanese Association for Digital Humanites	N/A	N/A
Humanistica	Humanistica, L'association francophone des humanités numériques/digitales	@HumanisticaDH	506

Table 5.1 shows that among the eight selected organisations, seven had Twitter accounts. On average, each account followed around 673 users.

The Japanese association did not have a Twitter account. However, it has an official Facebook group⁴⁷ with many active scholars posting news and communicating on a daily basis. At the time of writing, this group has 666 members. Although Facebook is

⁴⁷ Facebook group of JADH (Japanese Association for Digital Humanities), more information can be found: <https://www.facebook.com/groups/758758500904522/about/>

not the main focus and data source of this study, future work can be done to investigate the DH social network on other social media.

Specifically, this study used ‘Follow, search, and get users’⁴⁸ and ‘Get Tweet timelines’⁴⁹ APIs to extract data. Since Twitter made its API⁵⁰ open, simple and well-conceived, it not only benefits users and developers who analyse Twitter, but also has contributed to Twitter’s early success for bringing in new features and making a better user experience (Makice, 2009, p. 46).

5.1.1 The dataset

In total, 3,160 unique users who were followed by these accounts were selected. Among them, 4 users set their accounts and contents as private, and 2 users did not tweet anything. Thus, 3,154 user profiles along with around 6 million of their tweets were collected. The collected data covered the whole period on Twitter from 2006-03-21 when Twitter was launched online up to 2017-11-05 when the data was collected.

In the collected dataset, the number of users and tweets experienced significant but different growth patterns over time (see Figure 5.1 and Figure 5.2).

⁴⁸ More information can be found: <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-friends-list>

⁴⁹ More information can be found: https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline.html

⁵⁰ Twitter API reference index, <https://developer.twitter.com/en/docs/api-reference-index.html>

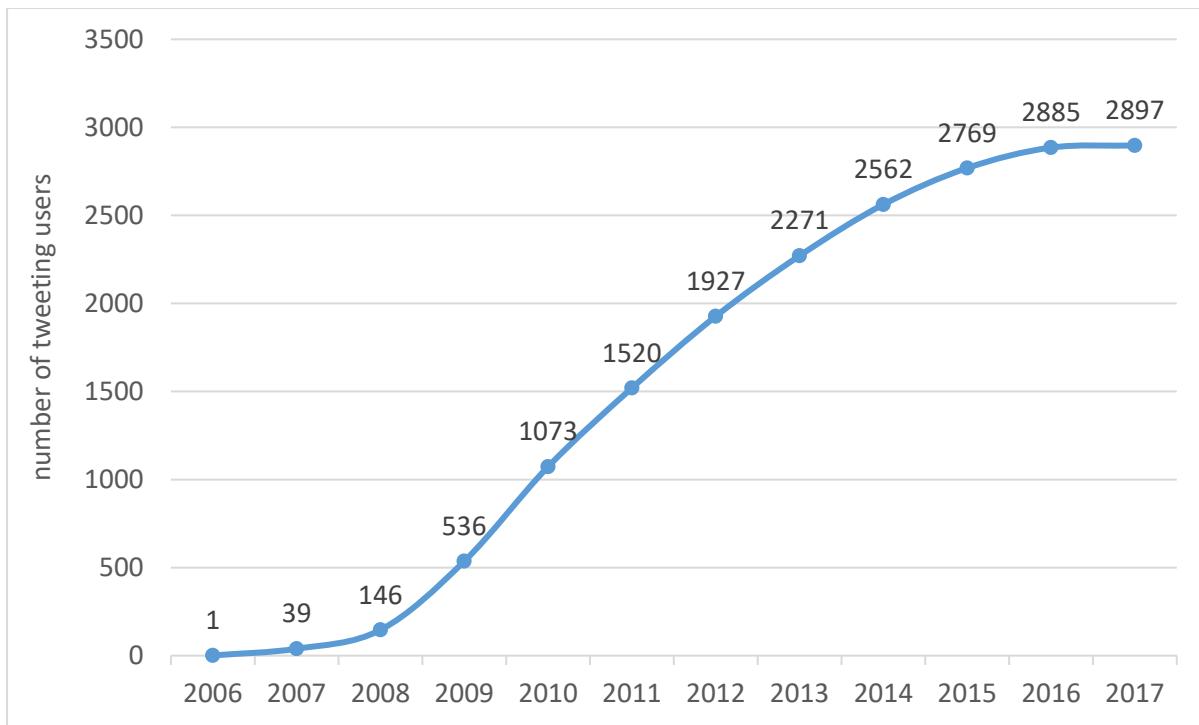


Figure 5.1: The number of tweeting DH users collected every year on Twitter.

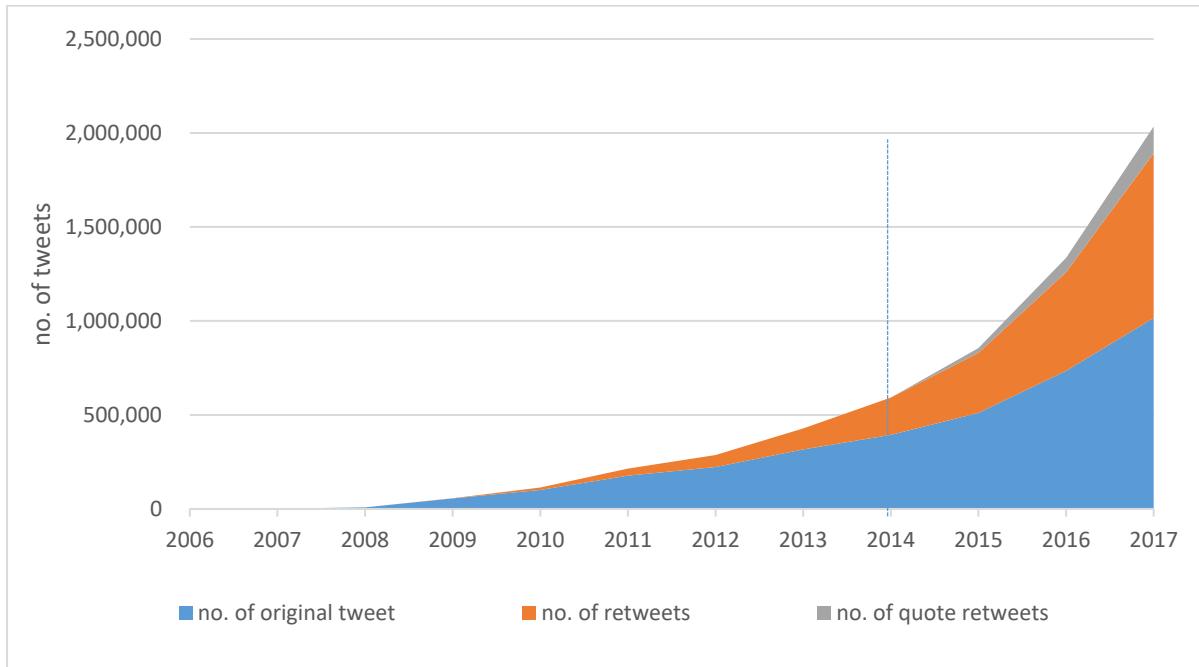


Figure 5.2: The number of tweets (including original tweets, retweets, quote retweets, and replies, etc.) from identifiable DH users collected each year.

Figure 5.1 shows a general growth of active users especially during the period from 2008 to 2014, whereas, more recently, from 2014 onwards, the number moved to a slower and more stable growth. In the area graph (Figure 5.2), however, the number

of tweets experienced an increase in growth in 2016-2017 after an initial steady rise from 2008 to 2014. This indicates that there has been a significant growth in Twitter activity in the last few years.

In particular, the number of retweets and quote retweets were growing rapidly, and such growth reveals a change in Twitter user activity. Figure 5.2 presents a turning point roughly around 2014 when retweet and quote retweets started to show a surge and reached about the same level in 2017. The more retweets and quote retweets (or comments), the more the users are aware of other users (and their tweets) whether inside or outside of their networks, and the more scholarly communications and information dissemination there might be. Indeed, as reviewed in previous studies (e.g., (Ross et al., 2011; Quan-Haase et al., 2015a)), scholarly discussion and information sharing were two of the main drivers that encouraged DH scholars to use Twitter, but how much do these two types of activities constitute towards their whole DH Twitter activities? How did they evolve over time?

With two analyses on the current dataset, this study is able to answer such questions by calculating the proportion of tweets that contain @ sign and URL over time.

Since 2009, when Honey and Herring firstly proposed the idea that the majority (around 90%) of tweets that contained @ sign were conversational (Honey and Herring, 2009), many studies used '@' as an indicator to identify the conversational tweets (Bruns, 2012; Inversini et al., 2015). Likewise, by calculating the proportion of tweets that contained at least one URL, one could also learn the proportion of information sharing activities on Twitter (Veletsianos and Kimmons, 2016). As shown below, Figure 5.3 is a line graph of the percentage of tweets that contained at least one @ sign (or mentioned other users) in the current dataset, while Figure 5.4 shows a line graph of the percentage of tweets that contained at least one URL.

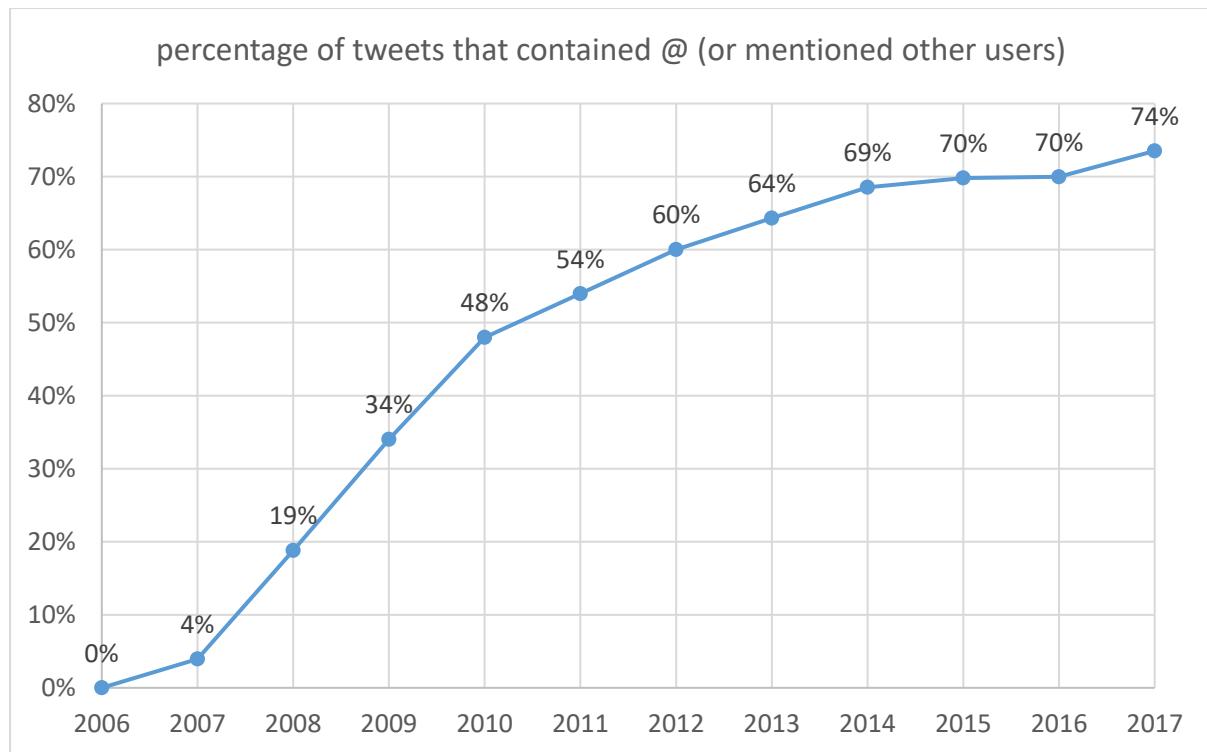


Figure 5.3: The relative percentage of tweets that contained at least one @ sign (or mentioned other users) in the current dataset.

In Figure 5.3, the proportion of conversational tweets in DH were increasing steadily. Especially after 2011, the majority (more than 50%) of the tweets were, according to Honey and Herring, conversational.

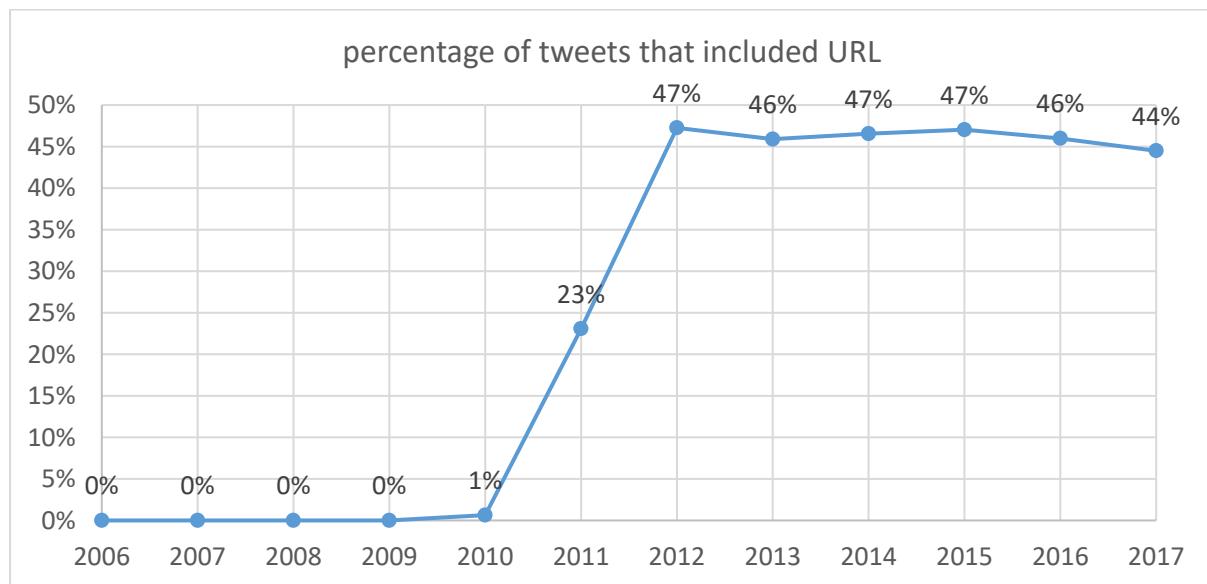


Figure 5.4: The relative percentage of tweets that contained URL in the current dataset.

Figure 5.4, however, shows that the proportion of informational tweets (containing at least one URL) in DH was vastly different to conversational tweets over time. Before 2010, there were hardly any tweets that included an URL, and during 2011-2012, the proportion suddenly shot up to around 47%. This might be because Twitter introduced the shortened URL format only in mid-2009 (Wortham, 2009). After 2012 and onwards, the percentage of link-sharing tweets remained at around 46% with a slight decreasing trend.

These two charts indicate how much communication and information sharing activities constituted towards the whole of DH Twitter activities, although not all tweets that contain @ handles are conversational (nor are all tweets contain URLs about sharing information). Nevertheless, the two graphs, at the same time, also raise new questions that seek to account for such differences between conversational tweets and informational tweets. These questions will be discussed in the following two network analyses at section 5.2 (Hashtag co-occurrence network) and section 5.3 (Co-retweet network).

5.1.2 Further data collection

Affiliated country information was gathered based on geographic locations that were provided by users in their profiles. However, the location field on Twitter allows users to have self-defined locations, for example, Toniesha L. Taylor (@DrTonieshaT)⁵¹ had 'Always Enjoying the Bounty!', Florence Chee (@cheeflo)⁵² had 'At work, rest, and play', and Ed Fay (@digitalfay)⁵³ had 'Work ⇔ not Work' as their locations. For such users, manual data cleaning was conducted to collect the user's affiliated country, if the affiliation was provided in the user profile. The default interface time zone was also used to help with the data cleaning. If the location or affiliation were not found, the

⁵¹ Toniesha L. Taylor, Assistant Professor of Communication and Interim Department Head in the Department Languages and Communication at Prairie View A&M University. <https://soundstudiesblog.com/toniesha-l-taylor/>

⁵² Florence Chee, Assistant Professor of Digital Communication and Director of the Social & Interactive Media Lab (SIMLab) at Loyola University Chicago. <http://simlabchicago.com/?p=127>

⁵³ Ed Fay, Associate Director at University of Southampton Library & Arts. <https://www.linkedin.com/in/digitalfay/?originalSubdomain=uk>

label ‘unknown’ was used. Similar to the bibliometric study, the country list on the United Nations website was also used for assigning the users’ affiliation.

Once the data was cleaned, the user country distribution based on the dataset can be seen in Figure 5.5.

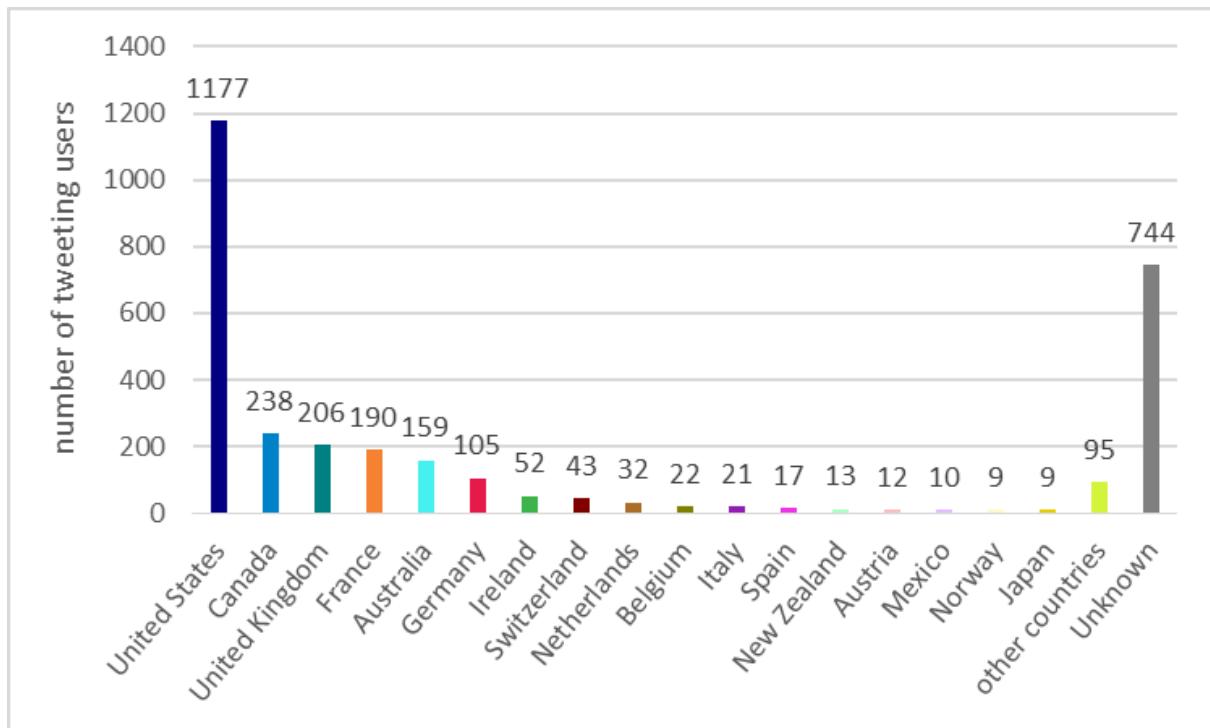


Figure 5.5: Bar chart of the regional distribution of collected tweeting users⁵⁴

Figure 5.5 shows an evident difference between countries. Among the 3,154 collected users, more than one-third of them are affiliated with the US and Canada. Apart from Australia, New Zealand, and Japan, the rest of the users are mostly affiliated with European countries, such as the UK, France, Germany, Switzerland, the Netherlands, Belgium, Italy and Spain. The distribution matches the time zone as well. There are 744 users (23.5%) who did not share their locations or affiliations on Twitter.

⁵⁴ Other countries include: Argentina (7), Greece (7), India (7), Luxembourg (7), Sweden (7), Poland (7), Portugal (7), China (6), Denmark (5), Czech Republic (4), Egypt (3), Finland (3), Lebanon (2), Brazil (2), Russia (2), Singapore (2), the United Arab Emirates (1), Malaysia (1), Hungary (1), Chile (1), Romania (1), Colombia (1), Cuba (1), Cyprus (1), Vietnam (1), Israel (1), Ukraine (1), Indonesia (1), Monaco (1), Nigeria (1), Dominica (1), South Korea (1), Venezuela (1)

Moreover, the data of Twitter interface language was collected and analysed (see Figure 5.6).

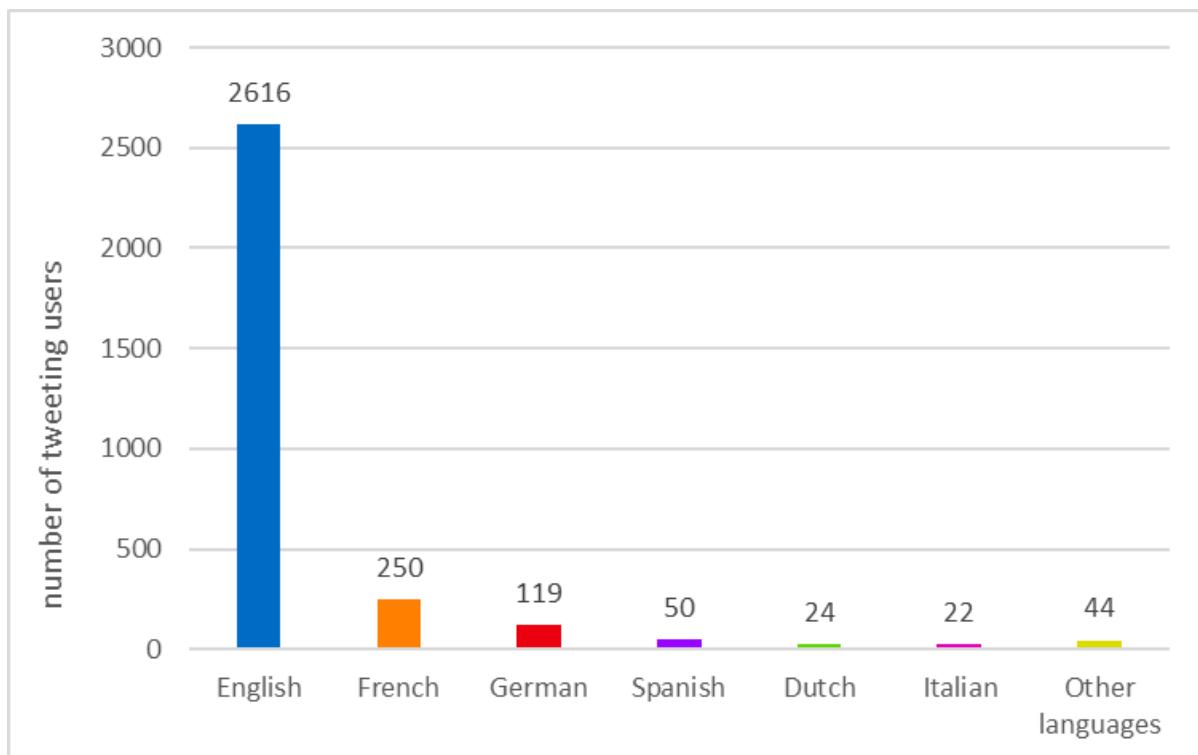


Figure 5.6: Bar chart of the interface language distribution of collected tweeting users⁵⁵

Given the country distribution, understandably, the majority of users were English speakers according to their Twitter interface language setting. However, this does not exclude the users whose first or main languages were in fact otherwise but chose English for working or communication convenience. As shown in Figure 5.6, we can also see other languages are used on the Twitter interface, such as French, German, Spanish, Dutch and Italian.

Despite the disadvantages discussed when using bio descriptions to identify the users, there is still much essential information that can be discovered from how digital humanists define or describe themselves on Twitter. The Figure 5.7 wordcloud was

⁵⁵ Other languages include: Portuguese (8), Japanese (8), Polish (5), Russian (3), Finnish (3), Catalan (3), Czech (3), Norwegian (3), Greek (2), Danish (2), Arabic (1), Swedish (1), Irish (1), Indonesian (1)

created based on the word frequencies of all the bios collected using wordle.net (top 5,000 words in corpus).



Figure 5.7: Top 5,000 words wordcloud based on twitter bios of selected users, constructed using wordle.net

As we can see from Figure 5.7, English is the dominant language appearing on the cloud. Among the top ranked words, most users had 'digital' (1,143) and 'humanities' (637) to describe themselves. Many users indicated their titles as 'professor' (227) or 'prof' (116), 'phd' (211), 'director' (180), 'historian' (157), 'librarian' (136), 'student' (119), and 'developer' (35). In terms of backgrounds, 'history' (393), 'university' (370), 'media' (205), 'library' (128) or 'libraries' (105), 'literature' (115), 'technology' (105), 'museum' (46) were brought up. In addition, many of them mentioned 'center' or 'centre' (100) or 'project' (98) that they might belong to, and some pointed to their affiliations by '@' (97) organisational Twitter handles on their bios. Some expressed their interests or values as 'open' (105), 'data' (162), 'views' (104) or 'opinions' (102), 'public' (86), 'world' (50), 'access' (43), 'enthusiast' (43), 'geek' (41), 'queer' (34), 'feminist' (32), 'food' (23), 'nerd' (20), 'gender' (20), etc. Others gave their regional information in their bios, such as 'Australian' or 'Australia' (52), 'London' (26), 'Canada' (25), 'European' (24), 'British' (23), 'French' (22), 'UK' (18), etc.

Additionally, we can also see the evidence of descriptions that were written in many other languages. When using the wordle.net online application, it automatically filters

the common English stop words (such as 'the', 'a', 'an', 'in' that a search engine has been programmed to ignore) but keeps most of the stop words of other languages. For instance, apart from 'numériques' (33) which means 'digital' in French, 'et' (122), 'en' (86), 'la' (74) might be French stop words that indicates the apparent use of French bios. While 'der' (21) and 'und' (17) are normally used in German language, and 'y' appeared to be the sign of Spanish descriptions. The word 'de' (215) seems to be used in the bios of Spanish, French, and Portuguese language speakers, and 'des' (67) and 'du' (40) might be for German and French descriptions.

Gender is another focus that attracts growing attention in DH. Hashtags of such discussions, such as #transformDH and #femDH, encourage participation of many DH users (Bailey et al., 2016). By using a similar name-gender assignment method in the bibliometric study (Larivière et al., 2013; Sugimoto et al., 2015), genders were assigned to the users as 'female', 'male', and 'unknown'. It is also noted that some people are gender diverse, but the sources for that information are very limited, so this study follows the previous gender category convention in (Rørstad and Aksnes, 2015, p. 321).

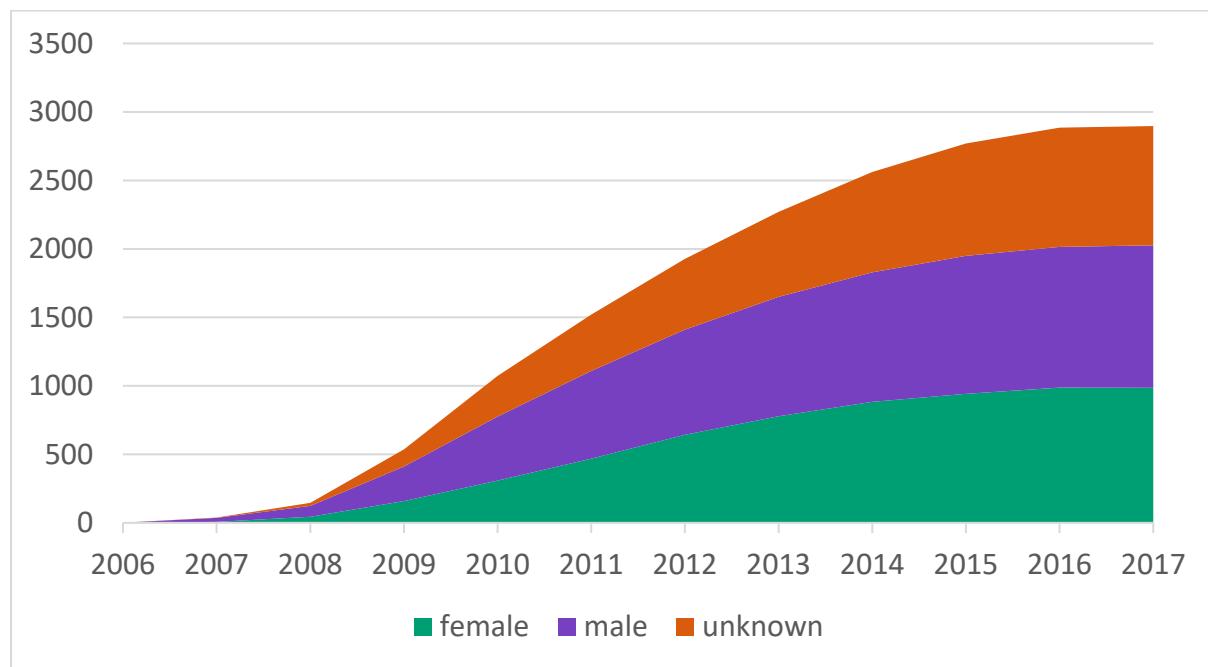


Figure 5.8: Number of female, male and unknown DH Twitter users each year.

As can be seen in the area graph (Figure 5.8), the users of the three gender categories are distributed evenly. There are in total 1,047 (33%) female, 1,098 (35%) male, and

1,009 (32%) unknown users. This gender dispersion is an improvement compared to previous DH gender studies or general Twitter gender surveys. For example, Fluharty found that there were 20% more male than female scholars on Twitter (Fluharty, 2010), and the figures were around the same (female/male was about 43/57) when Nielsen Mobile did the general survey across all users on Twitter (Nielsen Mobile, 2009).

From the graphs above, we can see a more diverse and gender-balanced DH community that contrasts with the bibliometric results. Further analysis will be discussed in the following two network analyses at section 5.2 (Hashtag co-occurrence network) and section 5.3 (Co-retweet network).

5.2 Hashtag co-occurrence network

The hashtag co-occurrence network is the third of the four networks that this thesis constructs. In general, the network analysis involves two parts, the node and the edge. Firstly, hashtags were extracted from individual tweets as nodes according to their number of occurrences (section 5.2.1). Then, the co-occurrence relations between each two hashtags were calculated as edges (or links) in order to discover the connections between the different hashtags (5.3.2). As mentioned, VOSviewer and Gephi were used for the network visualisation and centrality measures (section 5.2.3 and 5.2.4), and the time period (2006-2017) was split into individual years to study the hashtag topics longitudinally and to trace the development of DH topics on Twitter (section 5.2.5).

5.2.1 Node

There were 345,857 unique hashtags extracted from the current dataset, and they were used 3,031,157 times in total by these selected users. On average, one user tweeted (including retweeted, quote retweeted, and replied) 109.6 unique hashtags, and an individual hashtag was tweeted (and/or retweeted) 8.76 times. Figure 5.9 shows the number of unique hashtags used per year.

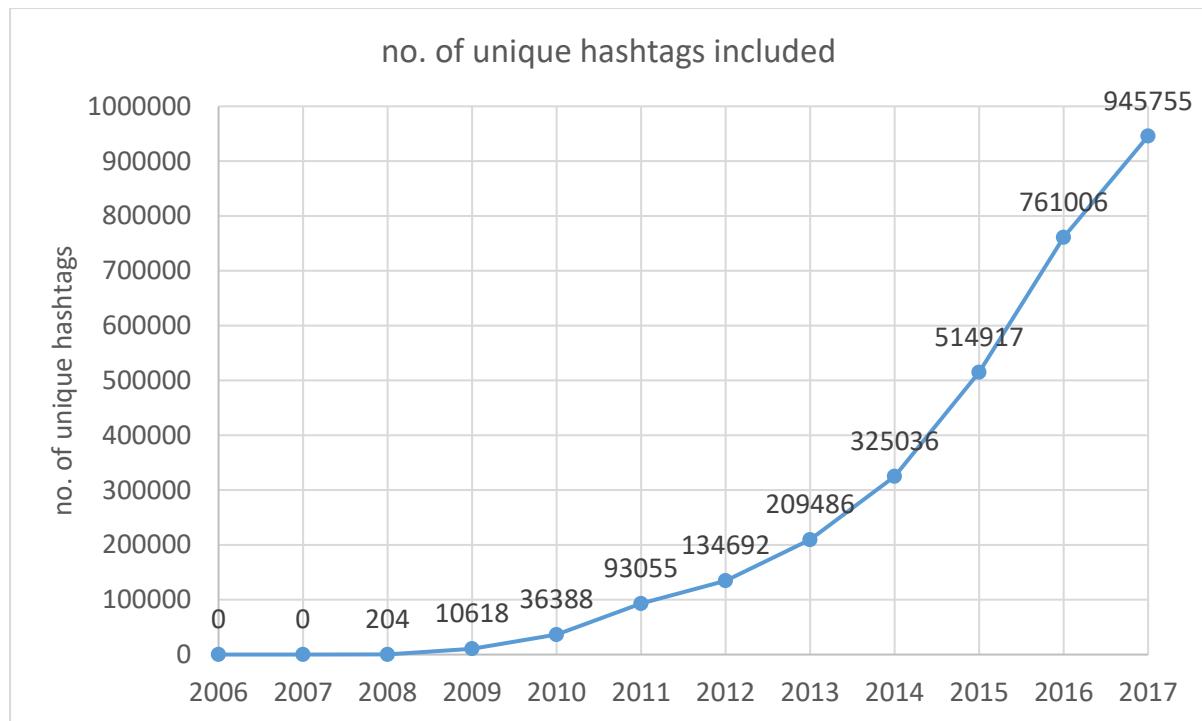


Figure 5.9: The number of unique hashtags used per year.

This line graph indicates a significant and continuous growth of hashtags used over time. Apart from the low numbers in early years (e.g., in 2006-2008) when users were learning to use the hashtag feature, the number of hashtags increased exponentially and became more diverse in later periods.

A node of a hashtag in this network analysis was weighted by counting the total number of occurrences in all tweets (including retweets, quote retweets, replies, etc.) that included this hashtag. For example, `#transformDH` occurred 3,239 times, and `#onthisday` occurred 7,691 times in the dataset, and therefore their sizes were 3,239 and 7,691 respectively. On rare occasions, the same hashtag might appear multiple times in one tweet, and, in that case, the total number of occurrences was counted towards the node size. For example, `#A` appeared 3 times in one tweet, then the total node size of `#A` increases by 3. However, in the current dataset, no hashtag appeared more than once in any individual tweet. Table 5.2 below, is an example of the top 30 hashtags ranked by total occurrences.

Table 5.2: The top 30 hashtags ranked by its total occurrences.

	hashtag	no. occurrences		hashtag	no. occurrences
1	#digitalhumanities	28,104	16	#bigdata	6,540
2	#dh	24,625	17	#data	6,016
3	#openaccess	14,799	18	#archaeology	6,008
4	#twitterstorians	12,935	19	#envhist	5,982
5	#dh2017	12,384	20	#libraries	5,920
6	#history	11,454	21	#dh2014	5,917
7	#opendata	10,569	22	#dhs2017	5,718
8	#humanities	10,187	23	#fb	5,443
9	#thatcamp	10,051	24	#iiif	5,350
10	#highered	9,200	25	#research	5,227
11	#dlfforum	8,641	26	#oa	5,218
12	#dh2016	8,128	27	#edtech	5,149
13	#onthisday	7,691	28	#digped	4,976
14	#archives	7,479	29	#dataviz	4,677
15	#otd	7,077	30	#dh2015	4,610

As can be seen from Table 5.2, the top 30 occurring hashtags are all related to DH but from different aspects.

Firstly, some hashtags are the disciplinary terminologies (e.g., #digitalhumanities, #dh) while others are DH-related events (e.g., #dh2017, #dh2016, #dh2015, #dh2014, #thatcamp, #dlfforum, #dhs2017). There are hashtags about open values (e.g., #openaccess, #oa, #opendata), and also there are topics about data analysis (e.g., #research, #bigdata, #data, #dataviz, #iiif). Various names of humanities disciplines are also used very frequently (e.g., #history, #libraries, #archives, #archaeology, #humanities), and particularly the topics related to history are very popular (e.g., #twitterstorians, #onthisday, #otd, #envhist, i.e., ‘environmental history’). Apart from these, education is also one of the main topics among the DH Twitter discussions, (e.g., #highered, #edtech, #digped, i.e., ‘digital pedagogy’). In addition, #fb appears 5,443 times, and it is used by people who use the automatic Twitter update application on Facebook where tweets ending with #fb are automatically exported to Facebook.

There are other popular topics in DH that are not related to DH research, for example #brexit occurred 2,536 times, #bikeschool 1,011 times and #fakenews 996 times.

These hashtags have not been removed from the current dataset as this study aims to explore a broad range of topics that DH users discuss and share on Twitter, and it is impractical to manually hand-pick and categorise the 345,857 unique hashtags within the time frame of this PhD study, but further future analysis is possible.

5.2.2 Edge

Similar to the bibliometric study, each edge on the hashtag co-occurrence network represents a link where any two hashtags appeared in the same tweet, and if such tweet was retweeted α times, then the value of their co-occurrence edge increases by α . For example, #a and #b were both included in 2 distinct original tweets, tweet 1 was retweeted 10 times by other users from the selected user pool, and tweet 2 was retweeted 5 times, then the co-occurrence of #a and #b is 17 (=2+10+5). The undirected edges of each pair of hashtags were calculated and an example of co-occurrence values is shown in below Table 5.3.

Table 5.3: Hashtag co-occurrence matrix of the top 7 tweeted hashtags with colour scale.

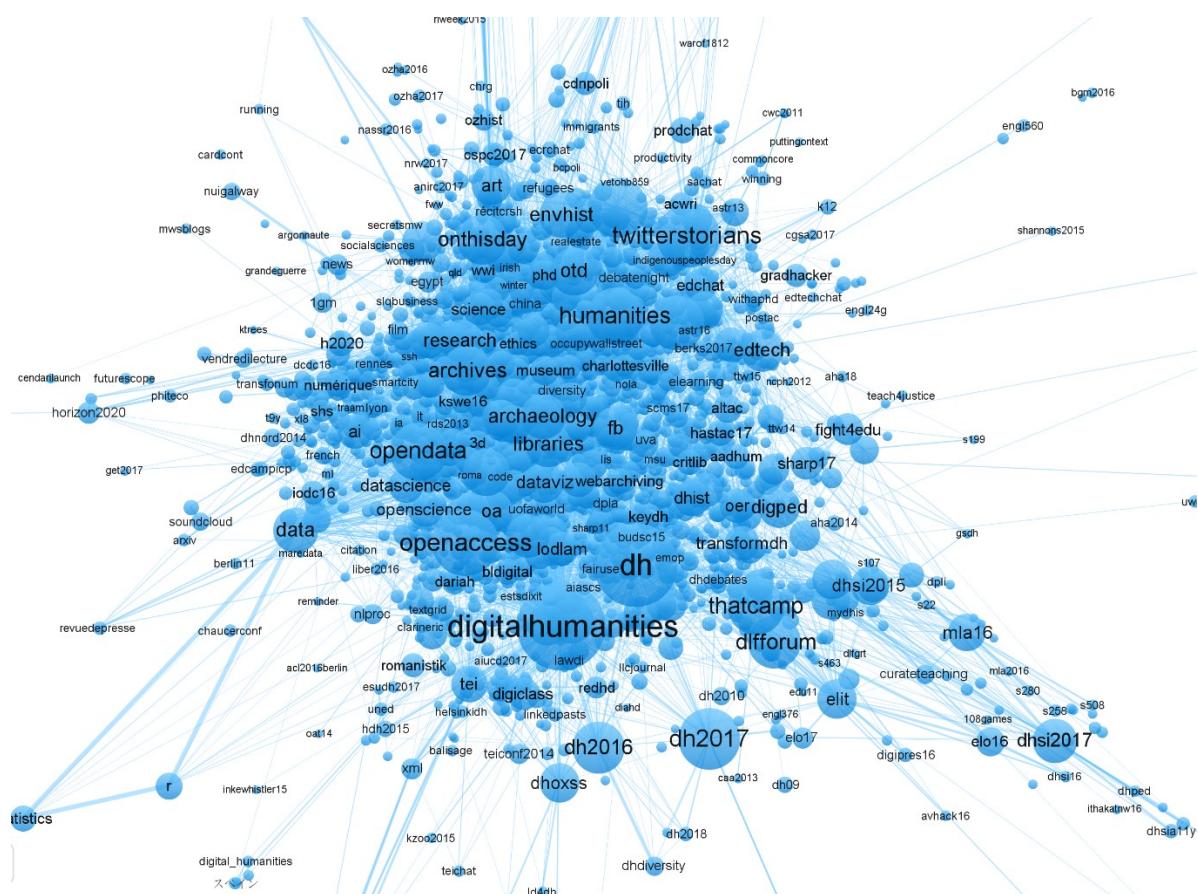
	#digital humanities	#dh	#openaccess	#twitter storians	#dh2017	#history	#opendata
#digital humanities	0	2251	198	383	214	156	83
#dh	2251	0	129	385	159	163	56
#openaccess	198	129	0	20	63	21	376
#twitter storians	383	385	20	0	2	951	0
#dh2017	214	159	63	2	0	1	0
#history	156	163	21	951	1	0	7
#opendata	83	56	376	0	0	7	0

As seen from Table 5.3 above, for example, #digitalhumanities had a very close relation with #dh (2,251 co-occurrences) while #dh2017 hardly had any connection with #history (1 co-occurrence) or #opendata (0 co-occurrence).

5.2.3 Network visualisation

By using VOSviewer, this study selected the top 3,018 hashtags (where the total occurrences were greater than or equal to 136 times) out of the 345,857 unique hashtags extracted from the dataset.

On the visualised network (Figure 5.10), around 98.34% of the hashtags (2,967) were connected to the network, which means that the majority of hashtags were co-included with other hashtags when DH users posted their tweets.⁵⁶ This percentage is much higher than the co-authorship network where only 19.54% of nodes were connected to the main network (see section 4.3.3).



⁵⁶ Because a hashtag is not case-sensitive, and having either uppercase or lowercase letters does not affect the hashtag value, this study converts all hashtags to lowercase for the convenience of data processing and hashtag identification. Whenever interpretation needs, however, capital letters are also introduced for better readability.

Figure 5.10: Hashtag co-occurrence network in DH, data extracted from Twitter, 2006-2017.

The difference in density in the networks might not be as surprising as we might expect, because previous studies also found the same. For example, after the analysis of around 9 million tweets, Türker and Sulak found that networks constructed by the co-hashtag method tend to have closer links and semantically related meanings, and thus it was common to form a dense and small world hashtag network (Türker and Sulak, 2017).

5.2.4 Centrality

Similar to previous networks, Gephi 0.9.2 was used to calculate the *betweenness centrality* (Brandes, 2001), and the complete table of results can be seen in Appendix D.

5.2.5 Longitudinal analysis

A synchronic view cannot reveal the dynamic topic changes along with Twitter application development. By visualising the hashtag network in individual years, this study effectively explores how the DH Twittersphere was formed and how DH Twitter history developed. This study visualises the networks of individual years during the 12-year period and the results will be shown and analysed in section 5.2.6.

5.2.6 Discussion and analysis

When looking at the Twitter subject network results, we find that Robertson's DH 'house structure' with 'many rooms' is no longer an appropriate description. The subject structure based on co-hashtag network does not have 'rooms' of disciplines, but rather 'gathering hall' of events, methods, languages, and values, with hardly any sign of 'rooms' with 'walls'.

As visualised in section 5.2.3 (Network visualisation), the hashtag co-occurrence network (Figure 5.11 below) is extremely dense, and one cannot find any apparent

clusters on such network.⁵⁷ Without drawing boxes, it is difficult to break down the overall image, locate positions and interpret the content. After further semantic analysis, we can see that the structure of DH topics on Twitter mainly consists of four sections that are closely connected and overlapping. Instead of using 'cluster' as the word to refer to an area on the network, 'section' has been used to better pinpoint and describe different semantic groups that are identified on the co-hashtag network.

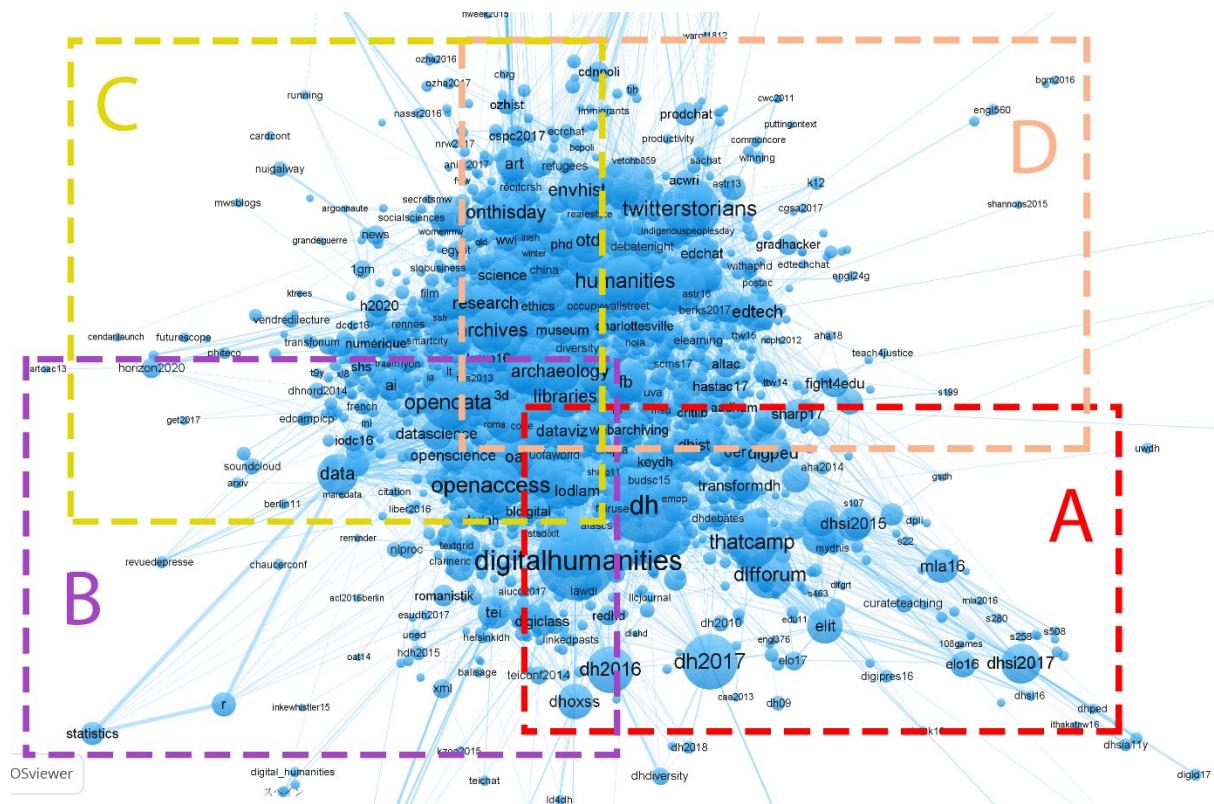


Figure 5.11: Hashtag co-occurrence network in DH, data extracted from Twitter, 2006-2017.

The number of nodes is distributed quite evenly at around 700-800 per section, although there are slight variations. The section A (DH and events) is the most recognisable group that attracts the majority of conference hashtags (e.g., ‘dh2017’, ‘dhs2017’, ‘mla16’, ‘thatcamp’), while section B (methods) includes more technical topics (e.g., ‘statistics’, ‘r’, ‘xml’, ‘tei’, ‘bigdata’). In section C (non-Anglophone DH), we can see many topics that are associated with non-Anglophone places, events, and

⁵⁷ Because in Figure 5.11, one cannot see any obvious separation or grouping, the density version of a 'heat map' is no longer useful to see the clustering, and this study chooses to present the 'normal' network version.

language keywords (e.g., ‘dhbenelux’, ‘dhberlin’, ‘humanitésnumériques’, ‘rome’, ‘méditerranéen’, ‘egypt’, ‘islam’, ‘africa’, ‘china’). Lastly, section D (history and other fields) is the largest section that contains topics related to history studies and other fields that are mostly humanities (e.g., ‘twitterstorians’, ‘history’, ‘onthisday’, ‘archives’, ‘museums’, ‘highered’). In addition, there is a very dense area at the centre of the network where the four sections overlap, such as ‘lodlam’ (Linked Open Data in Libraries, Archives, and Museums), ‘libraries’, ‘archaeology’, ‘dataviz’, ‘jobs’.

5.2.6.1 Section A – DH and events

Among all the hashtags, ‘digitalhumanities’ and ‘dh’ are the two most tweeted in the dataset (used 28,104 and 24,625 times, respectively), and centred around them are many DH-related events, such as DH2014-2017, THATCamp, DLFforum, DH Summer Institute, and MLA conferences (see Figure 5.12). Event hashtags are also among the popular hashtags, e.g., ‘DH2017’ (12,384 times). The majority of DH event hashtags can be found within this cluster. These event hashtags are distributed relatively distant to the whole network at the bottom right corner, and keywords such as ‘MLA16’ or ‘DHSI2017’ are located even further away (Figure 5.12). Compared to the distribution of other clusters (i.e., B, C, D) where nodes are positioned more closely, such detachment is not common, not to mention that event themes are often depicted as important factors of DH Twitter community formation.

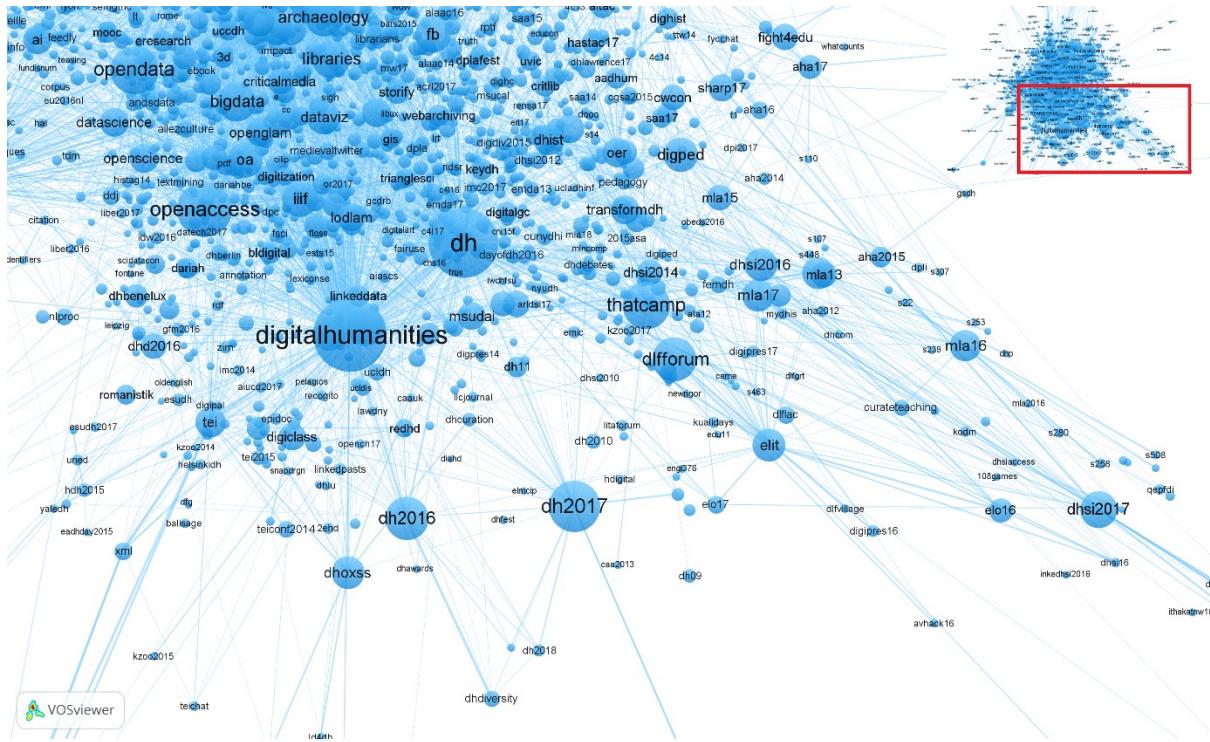


Figure 5.12: Bottom right part of the hashtag co-occurrence network in DH, data extracted from Twitter.

As shown in Figure 5.12, this detachment is unexpected when considering previous studies. The important role of Twitter as a ‘backchannel’ during DH events and conferences has often been mentioned as one of the features that characterised the DH Twitter community (e.g., Puschmann et al., 2011; Ross et al., 2011). Ross et al. suggested that a conference environment is an important encouragement to Twitter activities, and conference hashtags provide visible commentary and discussion that can form a reliable and searchable archive of events (Ross et al., 2011, p. 230). According to them, topics of events seem to be essential in the DH Twittersphere, and thus, it is surprising to find them on the margins of the network making distinct contrast with other closely grouped sections.

As shown on the network, DH events and conferences are not at the centre of Twitter discussions nor do they form the majority. The number of nodes in this cluster is less than a quarter of the total number of hashtags on the network, and it is also slightly less than the number of nodes in other sections. To explain such surprising results, we need to examine hashtag networks from a longitudinal perspective. As conducted in section 5.2.5 (Longitudinal analysis), hashtags related to DH conferences were

previously at the central position dominating the networks especially during the early period, but then gradually moved away to fringes after 2014. Networks of individual year are shown below (Figure 5.13 – Figure 5.21).

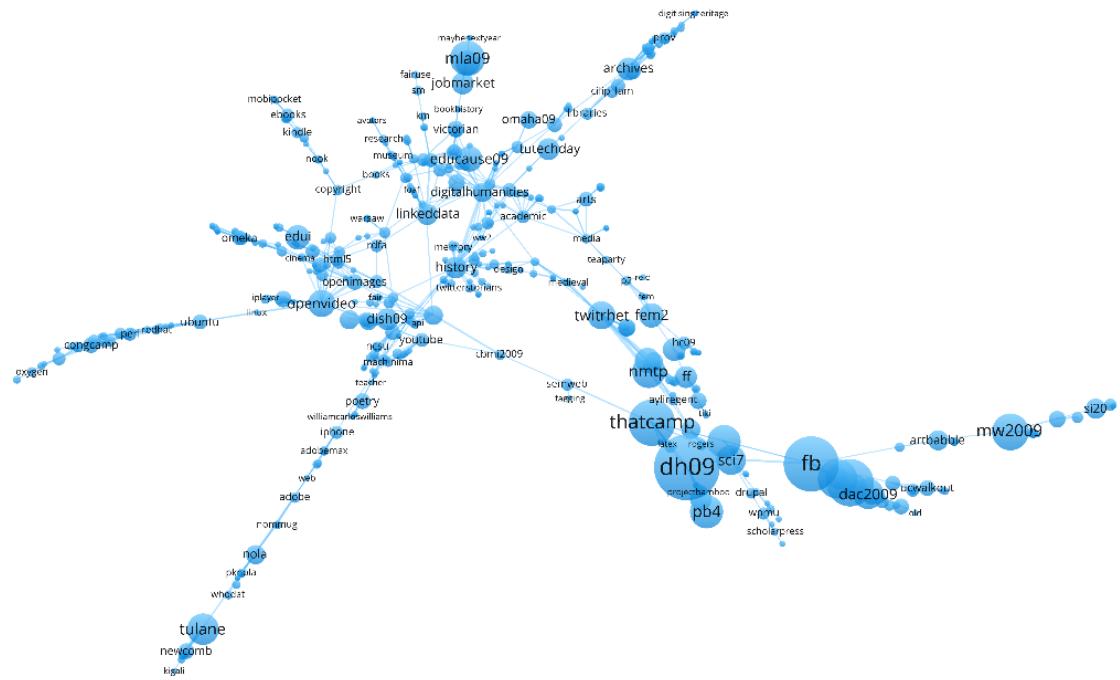


Figure 5.13: The DH hashtag co-occurrence network in 2009, data extracted from Twitter, 2006-2017.

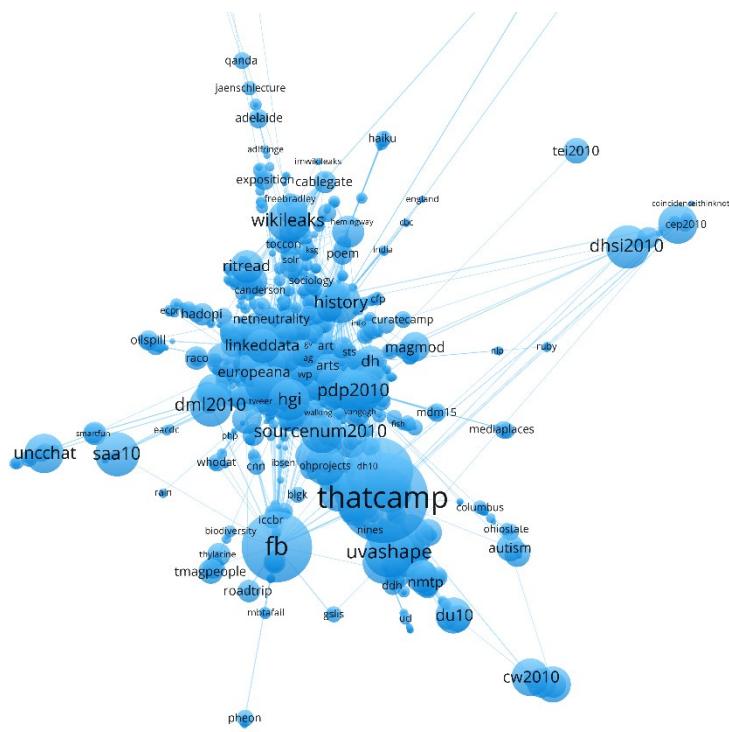


Figure 5.14: The DH hashtag co-occurrence network in 2010, data extracted from Twitter, 2006-2017.

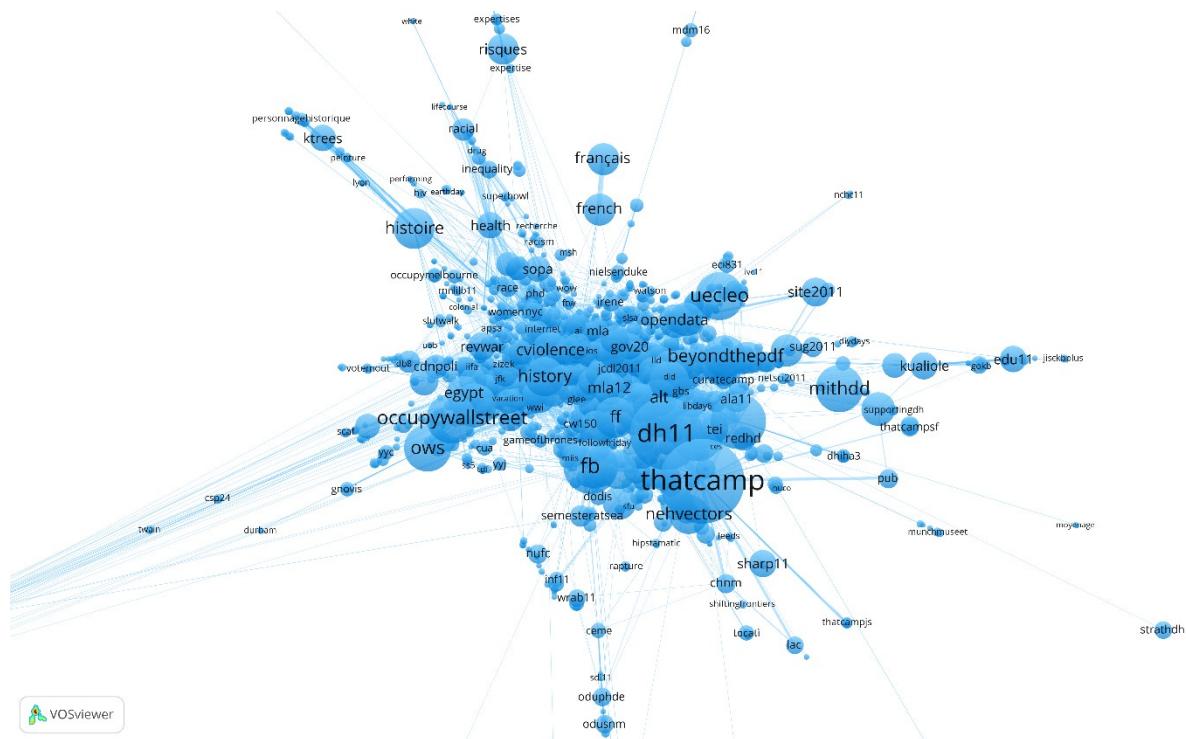


Figure 5.15: The DH hashtag co-occurrence network in 2011, data extracted from Twitter, 2006-2017.

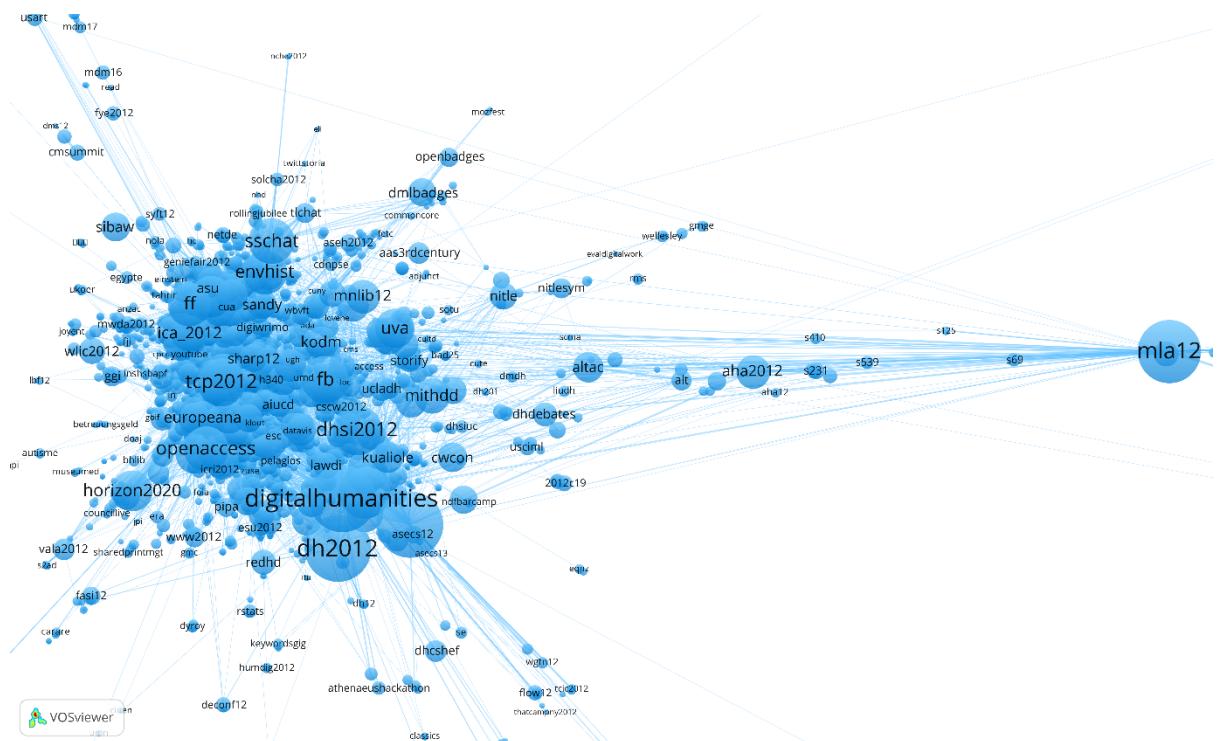


Figure 5.16: The DH hashtag co-occurrence network in 2012, data extracted from Twitter, 2006-2017.

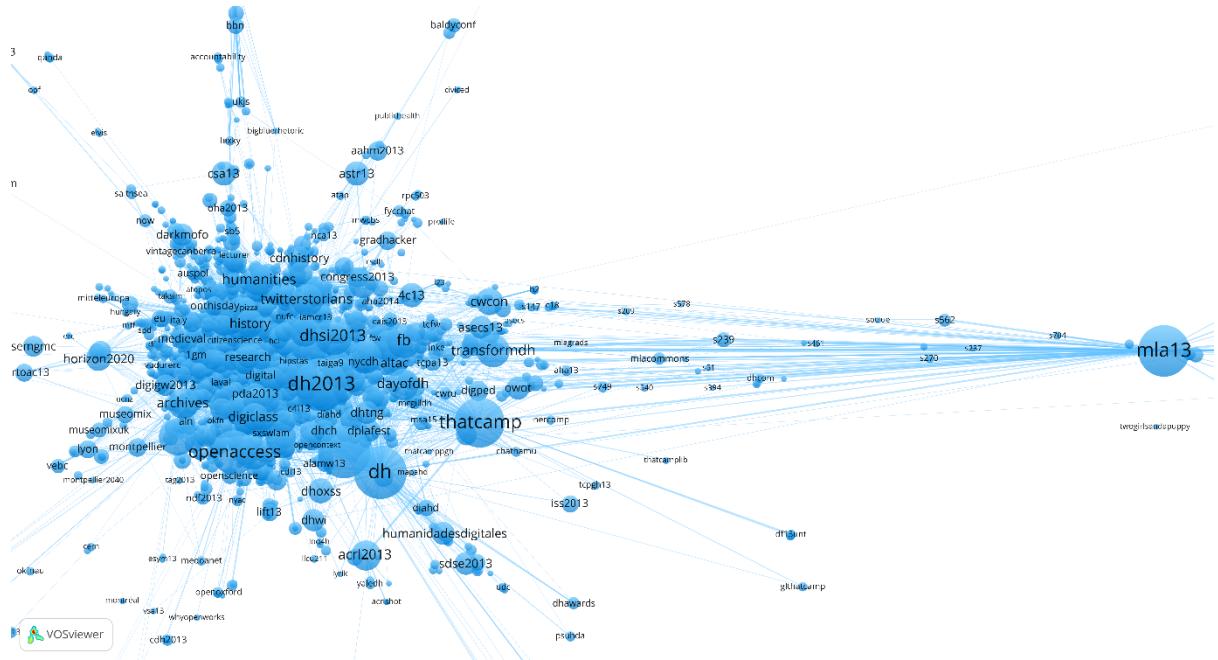


Figure 5.17: The DH hashtag co-occurrence network in 2013, data extracted from Twitter, 2006-2017.

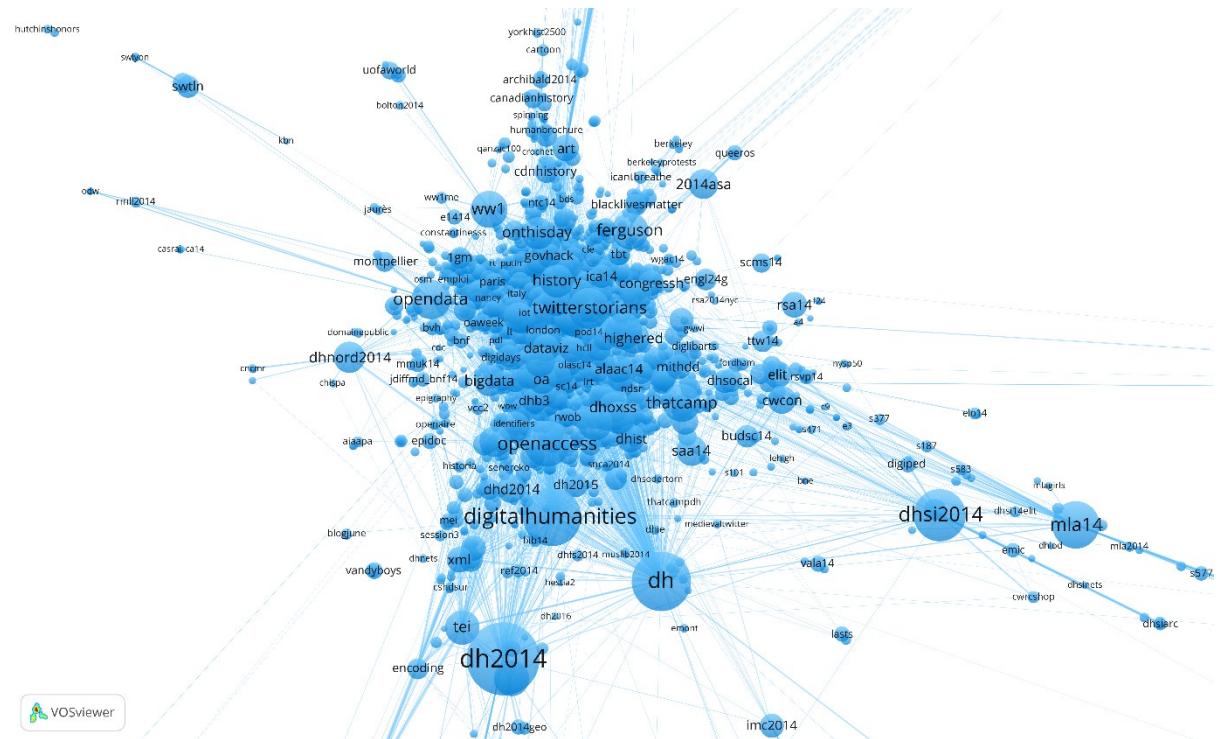


Figure 5.18: The DH hashtag co-occurrence network in 2014, data extracted from Twitter, 2006-2017.

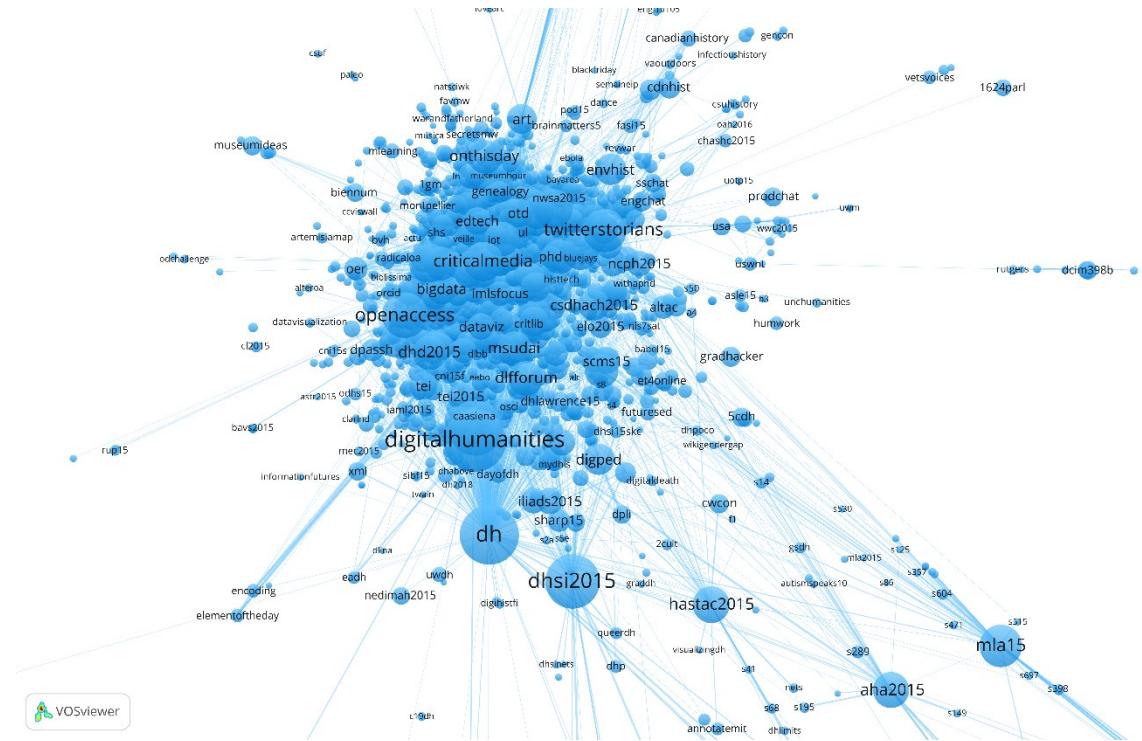


Figure 5.19: The DH hashtag co-occurrence network in 2015, data extracted from Twitter, 2006-2017.

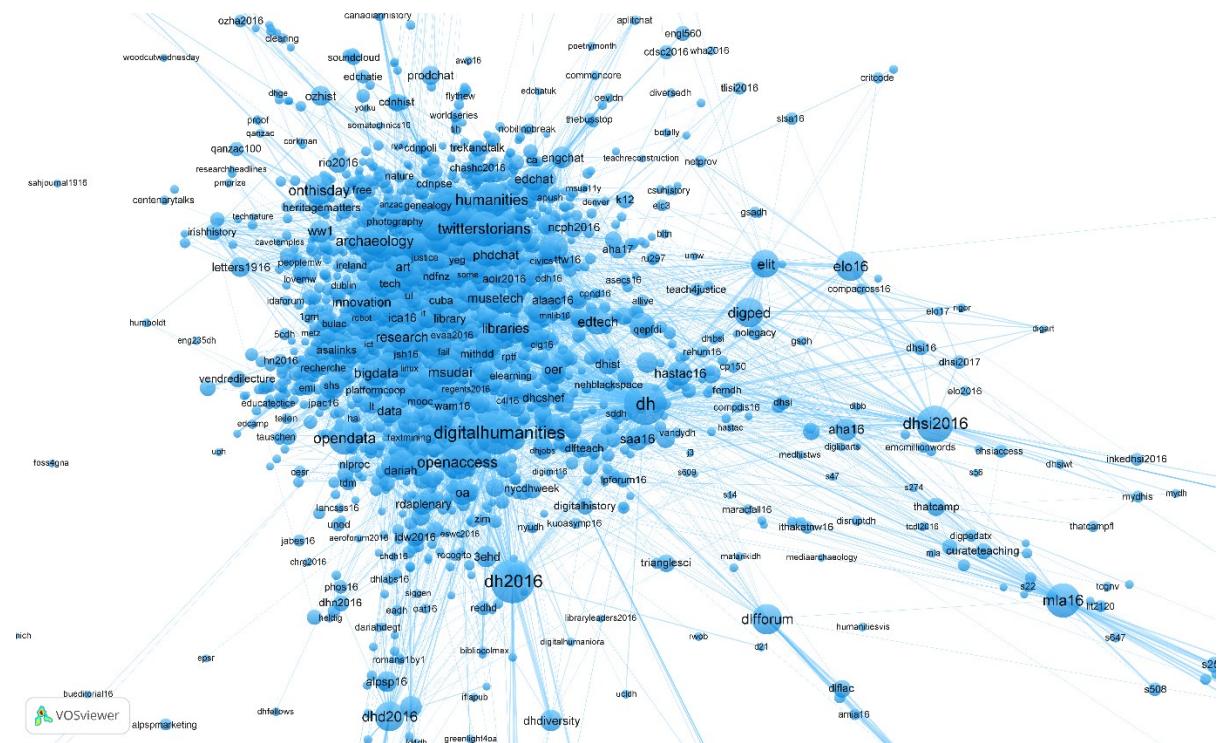


Figure 5.20: The DH hashtag co-occurrence network in 2016, data extracted from Twitter, 2006-2017.

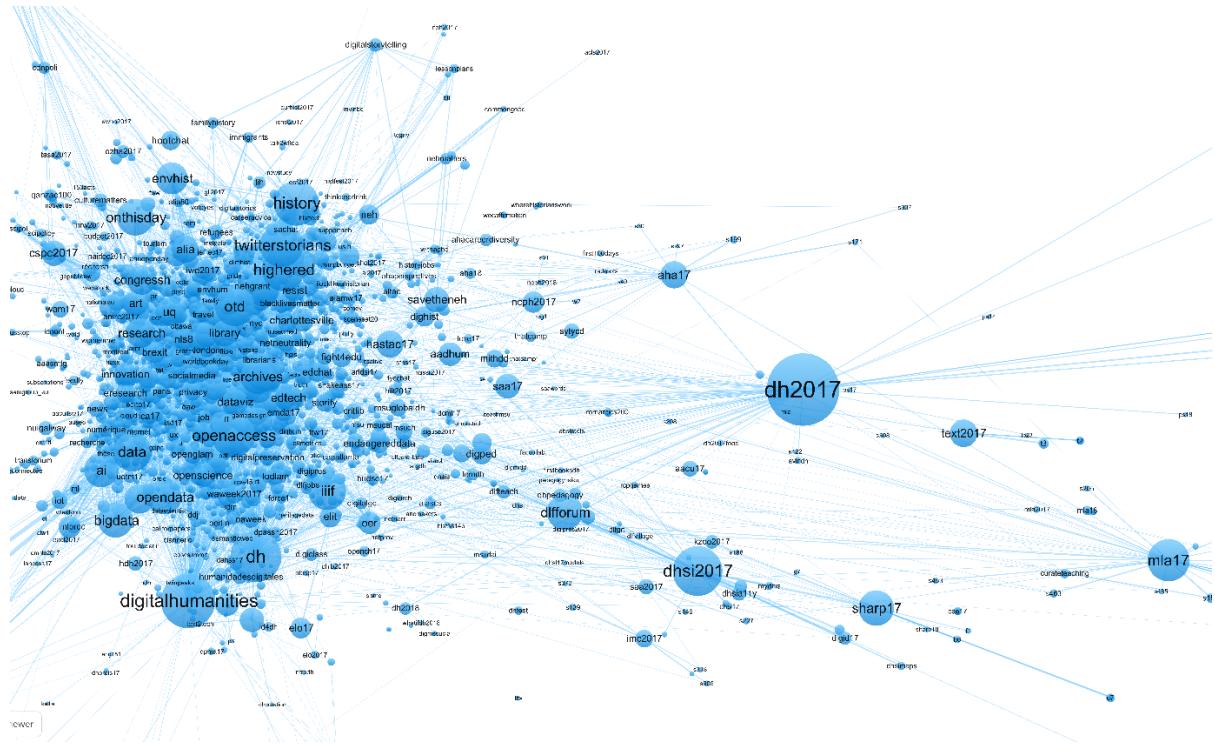


Figure 5.21: The DH hashtag co-occurrence network in 2017, data extracted from Twitter, 2006-2017.

In 2009 (Figure 5.13), the most popular hashtag was 'DH09' (tweeted 604 times in 2009) forming an obvious group of DH events on the right of the network with other hashtags, e.g., 'THATCamp' (240 times). Between 2010 to 2011 (Figure 5.14 and Figure 5.15), DH events continued to dominate the topics and remained as part of the central cluster, with the most used hashtags being 'THATCamp' (1,576 times) and 'DH2010' (729 times) in 2010, and 'THATCamp' (2,118 times), 'DH11' (1,479 times) and 'MLA11' (678 times) in 2011. During 2012 to 2013 (Figure 5.16 and Figure 5.17), MLA conference started to separate from the main network while other DH conferences still stayed inside and highly tweeted (e.g., in 2012, 'THATCamp' – 1,719, 'DH2011' – 1,711, 'MLA12' – 1,521; in 2013, 'THATCamp' – 1917, 'DH2013' – 2,223, 'MLA13' – 2,099). The year 2014 (Figure 5.18) is a turning point when most DH conferences started to separate from the main network, although conference hashtags were still among the most used ones (e.g., 'DH2014' – 5,686, 'DHSI2014' – 2,428, 'MLA2014' – 1,897). When separating from the main network, these event nodes did not move in the same direction and group together, instead, they started to spread and scatter across the lower right corner of the network.

During the next three years (2015-2017), most event hashtags remained outside of the main cluster, and some even moved further away (see Figure 5.19, Figure 5.20, Figure 5.21). For example, in 2016, 'DHSI2016' (4,274) and 'MLA16' (3,665) formed an area at the lower right corner, while 'DHd2016' (2,341) attracted some nodes and formed another cluster at the bottom of the network, although these two groups are small (Figure 5.20). In 2017, we can see that conferences moved further away from the centre forging individual small clusters around each conference hashtag like satellites (Figure 5.21), e.g., DH2017 (11,749), DHSI2017 (3,159), MLA17 (3,983).

From above, we learned that along with the growth of the DH Twitter community, the topic of DH events has been moving away from the centre of the DH Twittersphere since 2014, although the hashtags are still among the top-used. Whereas during 2009 to 2013, they were part of the main network and located centrally. Therefore, it is not difficult to understand why studies published before 2014 (e.g., (Puschmann et al., 2011; Ross et al., 2011; Ross, 2012) perceived conferences and events as the essential Twitter usage for DH people.

However, why did conferences start to separate from the main cluster in 2014? Is it because DH scholars' tweeting behaviour during conferences are in some way different from how they tweet about other topics (i.e., in section B, C, D) outside of conference time? If we take a look at the tweets during the DH conferences in the current dataset, we can find that compared to non-conference tweets, conference tweets had more mention (@) symbols and fewer external links, showing a stronger conversational usage during conference and informational usage during non-conference time (Table 5.4).

Table 5.4: Number of tweets with particular hashtag that included @ symbol and external link in the current dataset.

hashtag	no. tweets	no. tweets had @	% of tweet had @	no. tweets had link	% of tweets had link
#dh2016	8,128	6,087	74.89%	2,387	29.37%
#dh2017	12,384	9,329	75.33%	3,496	28.23%
#onthisday	7,691	3,697	48.07%	4,893	63.62%
#twitterstorian	12,935	8,457	65.38%	8,529	65.94%

As shown in Table 5.4, hashtag #DH2016 and #DH2017 in section A are more likely to be tweeted during a conference, while #onThisDay and #twitterstorians in section D are more likely to be tweeted during non-conference time. There are 74.89% of #DH2016 and 75.33% of #DH2017 tweets that included an @ symbol, while only 48.07% of #onThisDay and 65.38% of #twitterstorians had an @ symbol. As for external links, only 29.37% of #DH2016 and 28.23% of #DH2017 tweets included at least a link, but there are 63.62% of #onThisDay and 65.94% of #twitterstorians that had at least a link.

As mentioned earlier, users can ‘mention’ (‘at’ or ‘@’) other users by tweeting or replying ‘@’ with their user handles (or IDs). In this way, they can directly address other users, and the addressees will receive notifications when being mentioned. Many studies used ‘@’ as an indicator to identify the conversational tweets (Bruns, 2012; Inversini et al., 2015). For external links, on the other hand, by calculating the proportion of tweets that contained URLs, one could also learn the proportion of information sharing activities on Twitter (Veletsianos and Kimmons, 2016), although not all ‘@’ symbols are used for conversational purposes, and not all links indicate information dissemination.

Table 5.4 above demonstrated that DH scholars used Twitter for conversational purpose more often during conferences, while they posted non-conference tweets more frequently for information sharing. Moreover, before 2014, DH scholars on Twitter mostly tweeted about conferences (i.e., DH event was central 2009-2013). After 2014, the structure of topics became more diversified, and information sharing activities became more frequent, while conference topics moved to the margins but still remained popular.

These findings can thus clarify the question related to conferences in previous studies on DH Twitter usage. For example, both Ross and Holmberg’s studies found that compared to other disciplines, DH scholars used Twitter frequently for conversational purpose, especially during a conference period as an enabled backchannel (Ross et al., 2011; Holmberg and Thelwall, 2014). These studies were mostly done before 2014 when the DH Twitter topic structure was mainly centred on event hashtags, and therefore had more conversational contents. After 2014, non-conference hashtags formed a new cluster for the purpose of information sharing, and later became the main component of the DH hashtag co-occurrence network (Figure 5.10). It should be

noted that conversational and informational tweets do not exclude each other, and together, they develop a more vibrant academic Twitter community (Orduña-Malea et al., 2015; Kimmons and Veletsianos, 2016; Lim and Datta, 2016). Some scholars pointed out that along with the academic community development on Twitter, users would become more selective and experienced, and they would then perform both activities (Holmberg and Thelwall, 2014; Myers et al., 2014, p. 498).

In addition, there are other questions emerging from this cluster that require further analysis in future studies. For example, why did the year 2014 become the turning point when DH users became ‘more selective and experienced’ to perform both conversational and informational activities? We can see that users have a preference for using ‘digitalhumanities’ and ‘dh’, for the former is closer to the main network (section B, C, D) while the latter is closer to DH events – section A. Why would event attendees prefer to use the DH acronym? Among all the DH events, some are positioned relatively closer to the centre (e.g., ADHO conferences and THATCamp) while others are further from it (e.g., DH Summer Institute and MLA conferences). What are the forces that influenced such distribution? Why is MLA placed far away from the whole network? Is it the ‘signal shift in the MLA’s directive’ as mentioned by Pressman and Swanstrom in 2013 (Pressman and Swanstrom, 2013)? Also, in between section A (DH and events) and section D (History and other fields), ‘transformDH’ is placed very close to both, which indicates that it was tagged and used very often in both sections. This hashtag is a movement that started in 2011 (Kenny, 2015). It addresses a dynamic range of questions in DH (e.g., race, class, gender, disability, feminist, queer), and tries to shift the focus of DH from technical processes to social aspects (e.g., political, economic, personal, etc.,) (Bailey et al., 2016, p. 71). Other related hashtags can be found around these events, too, such as ‘FemDH’, ‘FemTechNet’, ‘DHpoco’ (i.e., DH post-colonial). Why do these DH values appear in the intersection of the DH event and History study but not in Non-Anglophone DH and data analysis area? Is it only the English-speaking DH community who pays attention to such values? This chapter will come back to discuss these questions to combine results in 5.2.6.3 (Section C – Non-Anglophone DH), 5.2.6.4 (Section D – History and other humanities fields) and 5.3.6 (Discussion and analysis).

5.2.6.2 Section B – Methods

Moving from right to left, the bottom left part of the network mainly shows the ‘technical’ side of DH, such as methods, data, the use and analysis of data, and various programming languages (see Figure 5.22 below).

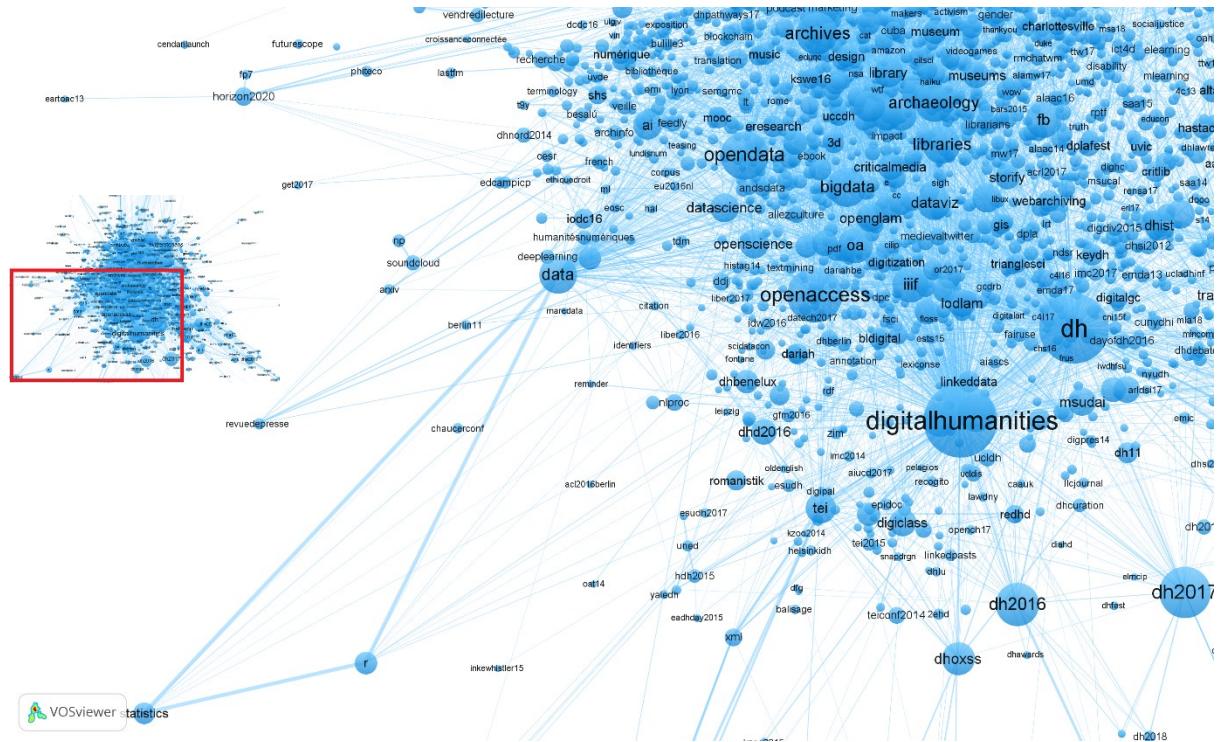


Figure 5.22: Bottom left part of the hashtag co-occurrence network in DH, data extracted from Twitter.

In Figure 5.22 above, many major hashtag nodes are related to ‘data’. On the top-left of ‘digitalhumanities’, there are nodes linked to ‘openaccess’ (‘OA’), such as ‘opendata’, ‘openscience’, while others are more connected to data research, such as ‘bigdata’, ‘AI’, ‘IIIF’, ‘DataViz’, ‘GIS’, ‘3D’, ‘digitization’, ‘MachineLearning’, ‘DeepLearning’. Further on the left, there are more technical terms, such as ‘TEI’, ‘XML’, ‘R’, and ‘statistics’. These technology terms and approaches often appear in discussions about DH practices in the current Twitter dataset. Although DH technologies vary and should not be seen as explicitly linked with the wider DH intellectual landscape (Kraus, 2013), as Skene argued, ‘in many ways, understanding digital humanities is easiest through grappling with its many methodologies’ (Skene, 2019). In general, these technical terms shown in Figure 5.22 cover most areas that we consider to be DH methods. For example, Burdick et al., proposed a list of DH method categories in the

chapter '*Emerging Methods and Genres*' of their book '*Digital_Humanities*' (Burdick et al., 2012). In their chapter, 15 types of DH methods were categorised and introduced, and we can find hashtags that are related to each type in section B as the following:

- 1) Enhanced Critical Curation (e.g., 'collection', 'museum')
- 2) Augmented Editions and Fluid Textuality (e.g., 'digitalEditions', 'TEI', 'XML')
- 3) Scale: The Law of Large Numbers (e.g., 'algorithm', 'bigdata')
- 4) Distant / Close, Macro / Micro, Surface / Depth (e.g., 'CloseReading', 'DistantReading')
- 5) Cultural Analytics, Aggregation, and Data-Mining (e.g., 'analytics', 'dataMining')
- 6) Visualization and Data Design (e.g., 'visualization', 'DataViz')
- 7) Locative Investigation and Thick Mapping (e.g., 'GIS', 'mapping')
- 8) The Animated Archive (e.g., 'archive', 'virtual')
- 9) Distributed Knowledge Production and Performative Access (e.g., 'crowdsourcing')
- 10) Humanities Gaming (e.g., 'gaming')
- 11) Code, Software, and Platform Studies (e.g., 'software')
- 12) Database Documentaries (e.g., 'narrative')
- 13) Repurposable Content and Remix Culture (e.g., 'remix', 'DigitalCulture', 'translation')
- 14) Pervasive Infrastructure (e.g., 'cloud', 'WebArchives')
- 15) Ubiquitous Scholarship (e.g., 'DigPublishing', 'CriticalMedia')

These hashtags can also fit in to other DH method classifications although the number of hashtags assigned to each category may vary significantly. The Oxford University DH programme uses high-level categories as following, and it can also accommodate many hashtags in cluster B (Hughes et al., 2015, p. 156):

- 1) communication and collaboration (e.g., 'crowdsourcing')
- 2) data analysis (e.g., 'visualization', 'DataViz', 'TextMining')

- 3) data capture (e.g., ‘digitization’, ‘transcribing’)
- 4) data publishing and dissemination (e.g., ‘publishing’, ‘OpenAccess’, ‘OA’)
- 5) data structuring and enhancement (e.g., ‘DataScience’, ‘IIIF’, ‘LinkedData’)
- 6) practice - led research (e.g., ‘practice’, ‘archiving’)
- 7) strategy and project management (e.g., ‘project’, ‘management’)

It is apparent that there are other DH method categories and classifications (e.g., Levenberg et al., 2018), and the concept of a ‘DH method’ itself is ever developing and changing.

Looking through the hashtags in section B and the different categories of DH methods, one might find that not all hashtags are about DH methods (e.g., many are about data and the use of data) and not all can be assigned into one of the categories above (some might belong to multiple ones or none). However, regardless of the classifications, we can see a very broad range of DH methodological discussions and topics in the visualised Twitter subject network. This wide range of topics matches some similar distributions at DH conferences that were analysed by Weingart (Weingart and Eichmann-Kalwara, 2017), e.g., ‘TextMining’, ‘GLAM’, ‘NLP’, ‘DataViz’, but the hashtags found in section B are not limited merely to the topics discovered from DH conferences. They are more diverse, and this is probably due to the different communication model and keyword/hashtag creation on Twitter compared to academic conferences.

For example, section B includes many hashtags expressing a strong value of openness (e.g., ‘OA’, ‘openaccess’, ‘opendata’, ‘openscience’, ‘openGLAM’), and also, there are many hashtags that reflect the existence of non-Anglophone DH communities (e.g., ‘humanitésnumériques’, ‘bibliothèque’, ‘DHBenelux’, ‘DARIAH-DE’, ‘DARIAH-BE’) which will be discussed in the next section. These hashtags are hardly seen among the DH conference keywords list that Weingart compiled (Weingart and Eichmann-Kalwara, 2017). Although these terms are not often used as research keywords in a broad sense, they are not surprising in the current dataset given DH’s encouragement towards open access and geo-lingual diversity, e.g., (Adema, 2014; Gil and Ortega, 2016). It is interesting, still, to see the keyword differences in DH tweets and DH conferences. The former represents what people talk about or discuss

on social media while the latter represents what people do or present at conferences. It seems that DH people talk more about OA and diversity but rarely research them or present on them at conferences. However, this is a finding that needs more careful mapping and comparing between the two sets of keywords and activities. Future study can be conducted to investigate the intellectual relationship between the DH Twittersphere and the published world.

Besides this, given that there is no clear clustering and the whole network itself is very dense, it is difficult to see apparent groups. This study, therefore, chooses to discuss the popularity of hashtags according to their relative distances from the centre of the network, i.e., the closer to the centre, the more frequently it appears with more other hashtags, and so arguably the more popular it is. This is a relatively new but well-grounded method to identify core-periphery structures, especially in densely connected network visualisations (Rombach et al., 2014)

Looking closely at the cluster B, we can generally organise its hashtags into two groups by distances to the centre. One group includes mostly data usage and data analysis terms, e.g., 'data', 'opendata', 'bigdata', 'openGLAM', 'IIIF', 'datascience', 'textmining', and they are located very close to, or even blended into the nodes at the centre, while the other includes various programming languages and approaches, e.g., 'R', 'statistics', 'XML', 'TEI', 'NLP', and they are positioned far from the main network with distant and loose connections. This clearly shows the preferences for topics towards the humanistic use, analysis and thinking of digital scholarship, although it does not mean that the DH Twitter community does not care about programming languages or statistics, because many of its methods are based on them. The visualisation is not surprising, and it confirms the importance and dominance of the humanistic elements in the DH intellectual map, just as Liu said in 'the meaning of the digital humanities':

In both their promise and their threat, the digital humanities serve as a shadow play for a future form of the humanities that wishes to include what contemporary society values about the digital without losing its soul to other domains of knowledge work that have gone digital to stake their claim to that society. (Liu, 2013, p. 410)

Nevertheless, we know that many DH methods are adopted and developed from other fields, e.g., social sciences, mathematics, and computer sciences (especially digital technologies and approaches). As DH continues to develop, it is now challenging to distinguish which methods are more ‘native’ to DH (Burdick et al., 2012, p. 30). By studying the hashtags in cluster B, especially two such groups by distances, however, we can still find something useful for identifying the core methods as well as the less popular and less discussed approaches in DH. Apparently on the network, the use and analysis of data are at the centre of the network linking many discussion topics while the programming languages are marginal topics that are not often involved in the DH Twittersphere.

It is admitted that organising the two groups of hashtags did not follow any standard classification system (apart from the core-periphery structure in networks), and it is debatable as many terms can be assigned to both groups that are not exclusive to one another. As many hashtags are not specific method names but rather types of data, standard, tool, and collection, grouping them by different distances on the network can assist further comparative studies of DH methods and classification and the studies of the meaning of digital humanities, bringing more quantitative evidence to the current discussions.

5.2.6.3 Section C – Non-Anglophone DH

Further above on the section B, we can see many topics that are related to events and themes in European countries and non-Anglophone languages. It is section C, and it is labelled as Non-Anglophone DH in this study (see Figure 5.23).

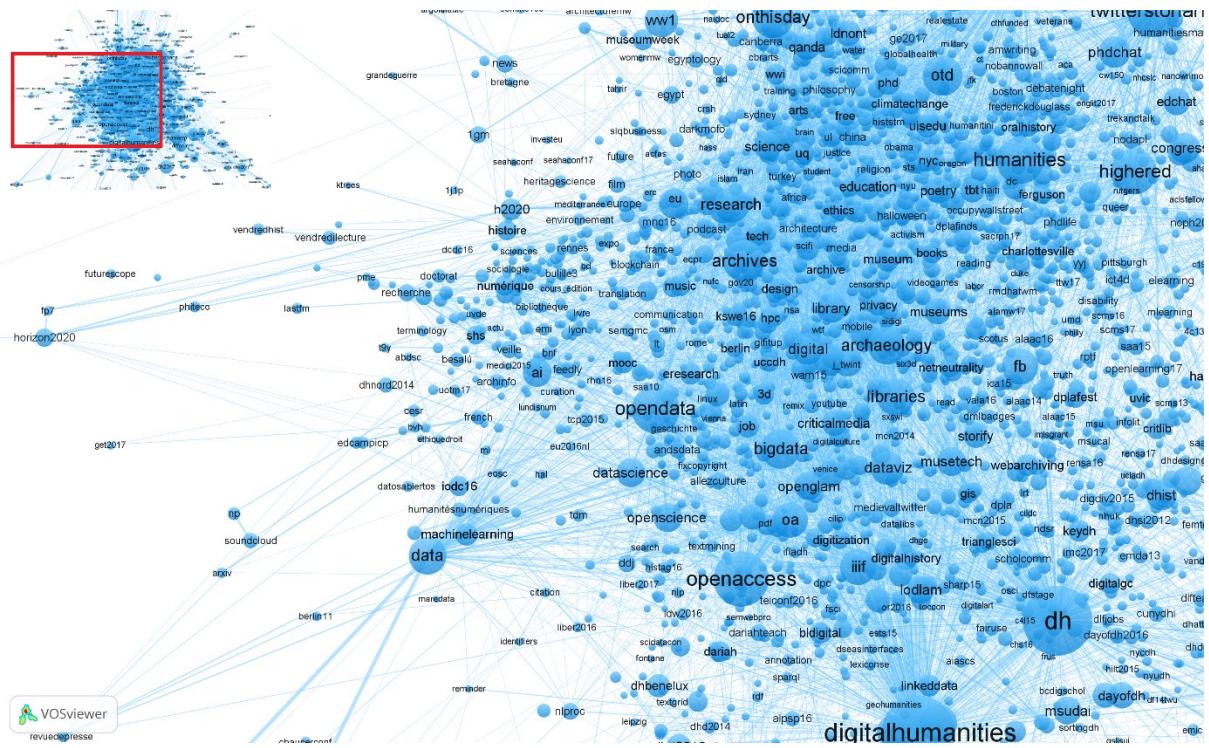


Figure 5.23: Left part of the hashtag co-occurrence network in DH, data extracted from Twitter.

In Figure 5.23, we can see EU-related hashtags, such as 'EU' (next to 'research'), 'europe', 'France', 'Berlin', 'Rome', 'Besalú', 'DHN'⁵⁸, and further on the left, there are 'Horizon2020' and 'H2020'⁵⁹. There are also topics that are related to other areas, such as 'méditerranéen', 'Egypt', 'Iran', 'Islam', 'Turkey', and 'Africa'.

According to an estimated calculation done by this study, among all the 2,105 hashtags shown in Figure 5.23, approximately 27.2% of the hashtags (573) are non-Anglophone keywords or related to non-Anglophone events and places. No other section in this network has such a high percentage.

As mentioned, these hashtags of a high level of geo-lingual diversity are mingled and associated with data and method topics in section B (Methods). For example, centred

⁵⁸ DHN is Digital Humanities in the Nordic countries, an associated organisation to EADH. It also organises an annual conference. More information can be found: <http://dig-hum-nord.eu/>.

⁵⁹ Horizon 2020 (or H2020) is the biggest EU Research and Innovation programme. More information can be found: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>.

around ‘openaccess’, ‘opendata’ and ‘openscience’ are many hashtags that are about German and Dutch language focused DH events, such as DHd2016-2017⁶⁰, ‘Leipzig’, ‘DHBENELUX’⁶¹, ‘DHBerlin’, ‘DARIAH-DE’, ‘DARIAH-BE’, and ‘romanistik’⁶². Also, placed around ‘data’ and ‘AI’ are many hashtags related to French-speaking events and topics, such as ‘humanitésnumériques’, ‘French’, ‘bibliothèque’, and ‘Lyon’. Around ‘science’ and ‘technology’, we can find hashtags of various places, e.g., ‘Africa’, ‘Turkey’, ‘Syria’, ‘Iraq’, ‘Cyprus’, ‘China’. Further down the network, there are Spanish language hashtags that are close to ‘TEI’, such as ‘HumanidadesDigitales’, ‘RedHD’, ‘2EHD’, ‘SeminarioTF’, ‘FilosofíaHD’. Although at the very left edge of the network, the Japanese hashtags are positioned close to ‘statistics’ and ‘R’, e.g., ‘あとで読む’, ‘スペイン’, ‘論文’.

Furthermore, in the previous section 4.2.6.3 (Cluster B – Computational linguistics), we have left a similar question to be investigated. When analysing nodes related to computational linguistics related nodes in the author co-citation network, we found that they had apparent connections with linguistic studies of non-Anglophone languages, especially German-Dutch linguistic studies (e.g., John Nerbonne, Peter Kleiweg, Wilbert Heeringa, Hans Goebl). At the same time, these computational linguists are also positioned very closely to many computer scientists and mathematicians, as mentioned, e.g., Joseph Kruskal, David Yarowsky, Kenneth W Church. It is worth examining the research interests of non-Anglophone scholars in DH, especially the relationship between the popular topics and the scholars’ geo-lingual backgrounds.

Meanwhile, it needs to be noted that this thesis is analysing data and results that are mostly published in English. Both the datasets analysed in the DH citation network and the DH Twitter hashtag network use English as their primary language. All the journals collected by this study are published in English. All the Twitter users collected in this study are followed by ADHO members’ Twitter accounts, and most of their

⁶⁰ DHd is an annual DH conference focused on DH topics in the German-speaking area. More information can be found from: <http://dhd2016.de/>.

⁶¹ DHBENELUX is an annual DH conference focused on DH topics in the Dutch-speaking area, such as Belgium, Luxembourg and the Netherlands. More information can be found from: <http://dhbenelux.org/about/>.

⁶² Romanistik is a hashtag related to German Romance studies. More information can be found from: <https://romanistik.de/>

tweets are in English, although the percentage of non-Anglophone tweets in the dataset has increased during its later period (e.g., from 2014 onwards). Consequently, should we narrow the broad assumption/question to: when scholars from non-Anglophone backgrounds use English to write DH papers or discuss DH on social media, they are more likely to talk about the technical aspect of DH, and why is that?

English is arguably the most used language in international academic communication in mostly all subjects and for all activities ranging from informal correspondence to formal collaboration, from online interaction to in-person meeting, and many more (e.g., conference, publishing, education). Widely used citation indices (e.g., SCI, SSCI, A&HCI) are found to be in favour of, and are promoting, English-language publications (Garfield, 1976). These indices contain an over-proportionate share of English-language publications, and 'English language journals of a certain quality tend to be included easier than journals in other languages of the same quality' (Sandelin and Sarafoglou, 2004, p. 7). Such 'linguistic imperialism', a term coined by Phillipson to question the spread of English in academia (Phillipson, 2012), is not only found among Anglophone countries, but also in other countries where English-language publications are evaluated as more prestigious and of higher quality, e.g., the Netherlands (Vandenbroucke, 1989, p. 1461), and Scandinavian countries (Nylen et al., 1994, p. 151). Ammon pointed out that 'many scientists, even those not particularly concerned about language questions, know about such evaluations intuitively and that this knowledge influences their own language choice' (Ammon, 2006, p. 15).

Scholars choose English as the language of publication rather than their own languages for many benefits, and the most promising ones are believed to be derived from the Anglophone academic market (Ammon, 2006, p. 7). This market helps them to reach a great number of readers, peers, potential job offers and research funding, and as it grows, thus, it also has the ability to maintain the quality of successful scholars and research standards.

Nevertheless, there are certain difficulties and challenges for non-Anglophone scholars to publish in English, especially for those languages that are linguistically distant to English (e.g., Chinese, Japanese) and do not have their roots in Indo-European or Romance languages. It requires great effort to learn and improve their

English language skills as well as costs for producing English texts, and even with well-copyedited English content, some information might be lost or become less satisfactory (especially some linguistic specific concepts that are often found in the Humanities debates). Such difficulties are also found in academic communities whose own languages are relatively closer to English, and as Ammon pointed that:

Only one in 20 German professors of English (!) chosen at random claimed to be able to produce publishable texts in English all by themselves; the others, the great majority, confessed the necessity of native-speaker assistance. In light of such findings, one can imagine the enormous linguistic difficulties of producing publishable texts for scientists whose language is linguistically more distant from English and whose own field is not English itself. (Ammon, 2006, p. 13)

For the Humanities and Social Sciences, in many cases, the knowledge is heavily dependent on their native languages, as Ammon gives the example:

e.g., the philosophy of Georg W.F. Hegel, who uses the three different meanings of the German verb **aufheben** ‘to raise’, ‘to abolish’ and ‘to preserve’ to develop his theory of dialectics; similarly, with other vocabulary, Martin Heidegger and numerous other thinkers (Ammon, 2006, p. 17)

When writing more humanistic texts, non-Anglophone scholars have a higher chance of using incorrect English terminologies, or receiving erroneous corrections from native copyeditors who are not familiar with the field knowledge (Matarese, 2013, p. 258). Consequently, they may end up with many incorrect expressions without being aware of them, not to mention maintaining the quality of their publications.

Therefore, although it is not the only reason, getting their work published in English with high quality is relatively easier through the technical point of view. Some studies found that technical languages in science communication are usually more formalised, functional, and have widely-used English terms, which makes it easier to handle for foreign-language speakers (Ammon, 2006, p. 4).⁶³

⁶³ In this study, the term ‘science’ (or ‘scientific’) is used to include all subjects in the academia, but sometimes, it is also used to be in opposition to certain disciplines like the humanities, and when necessary this will be clarified in the context.

On the other hand, because humanistic topics are often dependent on their native languages, many of the texts might be published in their native language publications and platforms, and thus we cannot see them in English publications, e.g., ADHO journals and Twittersphere. This might also be one of the reasons that non-Anglophone content in the current dataset is mostly associated with technical topics. It is worth investigating further to reveal the geo-lingual diversity in DH and the specific language community representation. In the coming sections, this study will combine the language and affiliated country data to continue discussing the different topical distributions for diverse lingual communities and the possible reasons for them.

5.2.6.4 Section D – History and other humanities fields

Finally, the last part of the network is mostly about various humanities disciplines that are involved in the DH Twitter discussions (especially, history, education and politics related topics), and they are positioned at the top of the whole network occupying the largest part (see Figure 5.24). Such humanistic topic distribution is quite different to the intellectual structure found in the co-citation network, and this section will compare and analyse the differences. As almost all hashtags in this section are in English, this study also explores the possible connection between the humanistic topics and the Twitter preferences of the Anglophone communities.

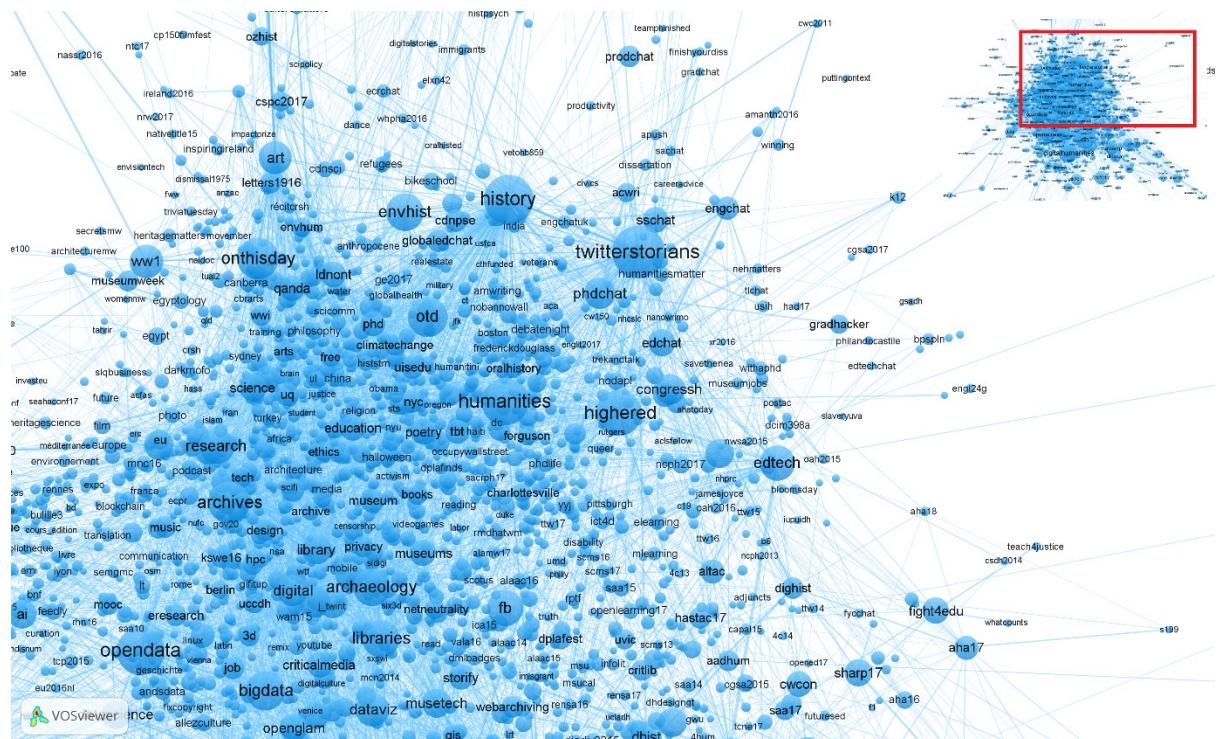


Figure 5.24: Top right part of the hashtag co-occurrence network in DH, data extracted from Twitter.

We can see in Figure 5.24 that many hashtags are related to the humanities fields, which are the main research subjects in DH. Although this section includes hashtags related to most of the humanities disciplines (Stanford University, 2019), they weigh differently. ‘History’, ‘libraries’, ‘archives’, ‘archaeology’, ‘museum’, ‘arts’, ‘poetry’ are obvious nodes, while ‘philosophy’, ‘religion’, ‘music’, and ‘linguistics’ are relatively small and difficult to find. ‘History’ related hashtags, in particular, account for a large proportion of the network, such as ‘twitterstorians’, ‘onthisday’ (or ‘otd’), ‘envhist’ (i.e., ‘environmental history’). Education is also one of the main fields, such as ‘highered’, ‘edtech’, ‘phd’, ‘edchat’, ‘digped’ (i.e. ‘digital pedagogy’). We can also see many hashtags related to politics and news (especially news in Anglophone countries) spread across this part of the network, although some might not be directly related to DH research, such as ‘climatechange’, ‘Obama’, ‘Charlottesville’, ‘Boston’, ‘Ferguson’, ‘NYC’, ‘Oregon’, ‘Pittsburgh’, ‘refugees’, and ‘Halloween’.

Given that it is the largest section in the network, we can generally conclude that DH people mainly talk about the humanities on Twitter, and they tend to favour history, education and politics, in particular, or at least the Anglophone DH people. Based on the most-used 100 hashtags in the current dataset, this study detects 12 types of humanities fields based on the candidate’s analysis of DH hashtags. A margin of error is acknowledged (e.g., many hashtags belong to multiple types, no standard taxonomy), and further evaluation and disciplinary classification might be needed to continue to investigate this question. Table 5.5 below shows the 12 types of fields, hashtag numbers, total occurrences, occurrence percentages, and examples for each type.

Table 5.5: Types of the top 100 used Twitter hashtags in the current dataset

type	no. hashtag	occurrences	occurrence percentage	hashtag example
history	20	81,218	31.16%	twitterstorians, history, onthisday, otd, envhist, publichistory, cdnhistory, digitalhistory
education	19	55,072	21.13%	highered, edtech, digped, phdchat, oer, edchat, education, fight4edu
politics	18	33,952	13.03%	congressh, transformdh, ldnont, brexit, blacklivesmatter, womensmarch, nyc, trump
museum	7	13,191	5.06%	museums, digitalpreservation, museum, museumweek, askacurator
library	5	14,114	5.41%	libraries, library, dplafest, critlib
art	3	7,460	2.86%	art, design, arts
poetry	2	2,747	1.05%	poetry
humanities	1	10,187	3.91%	humanities
archaeology	1	6,008	2.30%	archaeology
music	1	1,617	0.62%	music
literature	1	1,616	0.62%	literature
other	22	33,485	12.85%	qanda, cwcon, storify, tbt, socialmedia, engchat, prodchat, twitter, culture, tbreaktweets, acwri

According to the table, we can see that three types of topics (history, education, politics) account for more than 65% of hashtag usage. This topical distribution is different to the previous studies, although there are not many. As previously reviewed in chapter 2, based on data from 2009 to 2012, Moravec found that the main subject that DH scholars discussed on Twitter was the central values of the field (e.g., ‘collaboration’, ‘diversity’, ‘encourage’, and ‘support’). Moravec also found that the DH community on Twitter was ‘very quiet about teaching’, while this is not the case shown on the network here where 21.13% of the top 100 hashtags are related to education. By comparing this with the previous studies, it seems that the discussion topics on Twitter have been changing over time and also as more new users have joined the discussions.

While such distribution of topics might not be a surprising result to many DH scholars on Twitter, when compare to the bibliometric structure of topics visualised in the author co-citation network (discussed in section 4.2.6), one can see significant differences. The co-citation network contains four popular topics – general historical literacy and information science (cluster A), computational linguistics (cluster B), English studies

(cluster C), and early pioneers (cluster D). Although the cluster of early pioneers does fit into the history theme, there are hardly any apparent history-related areas on the co-citation network based on DH journal publications 1966 – 2017, not to mention education and politics.

As mentioned earlier, there is limited overlap between the two groups of people collected from publications and Twitter, and this might contribute to the difference in topics. However, there are many people who appear in both networks. In this dataset, there are 367 unique author names (out of 3,382) from the collected publication data that match exactly the same usernames collected from Twitter (3,154 users in total). It also should be noted that many Twitter users choose not to use their precise name as their username, and this number (i.e., 367) does not include them. The actual number, therefore, should be much greater. In other words, there should be much more than an 11% overlap between the two groups of people. If all else is equal (*ceteris paribus*), technically there may be a much greater than 11% of overlap of topics in these two networks, but one does not see such correlation.

A significant possible reason would be the different time periods for the data collection. The bibliometric data was gathered from publications between 1966 to 2017 while the Twitter data was from postings between 2006 to 2017. Strictly speaking then, we are comparing the formal knowledge of the past five decades with the informal knowledge of the past decade. It is thus, not surprising to find an obvious difference in the topical distribution. In addition, as mentioned, the two publishing channels are very different, with one being well-written and carefully peer-reviewed and the other being abbreviated and timely posted.

Nevertheless, comparing topics between bibliometric and Twitter networks is a meaningful task, and it can provide new insights to our understanding of the field's development. This is particularly the case as some popular topics in DH first started to gain attention on Twitter, and were later published in journals and books (e.g., Bailey et al., 2016; Ahmed et al., 2018). More recently, as user-generated content on social media has interwoven with traditional news sources and mainstream media, Twitter has been considered to be a valuable historical source on 'history-as-it-happen' and has attracted a lot of research attention (Bruns and Weller, 2016, p. 183). Bruns and Weller pointed out in their article titled '*Twitter as a first draft of the present: and the*

challenges of preserving it for the future' that Twitter was not only of interest for contemporary journalism but also for future historians on user interests and future development (Bruns and Weller, 2016, pp. 183–184). It not only archives what happened in the past, but also implies popular topics for further discussion. Although it is still questionable if the DH distribution of topics for publications will develop eventually to be similar to the Twitter distribution that we see today, it is worth examining such tendency through a longitudinal perspective.

In order to study the development of Twitter topics over the past decade, this research chooses to look into the three most popular types of hashtags that were mentioned earlier (i.e., history, education, and politics). Note that the current data is incomplete for the years of 2006 to 2008, and many of the hashtags mentioned had not yet been introduced. However, it does not skew the results or introduce any bias that the candidate is aware of. This study uncovers how these topics were formed and developed and sheds light on future possible development.

Firstly, Figure 5.25 below shows the annual percentages of the most popular history-related hashtags. To calculate the percentage, the occurrences of each hashtag in each year have been counted first. For example, the hashtag 'twitterstorians' occurred 12,935 during the whole period, 3 times in 2009 and 4,222 times in 2017 respectively. Thus, the occurrence percentage of 'twitterstorians' in 2009 is 0.02% (3/12,935), and in 2017 it is 32.64% (4,222/12,935). The sum of all the percentages of each hashtag should be 100%, and by doing that, we can compare hashtag usage of different popularities at the same scale.

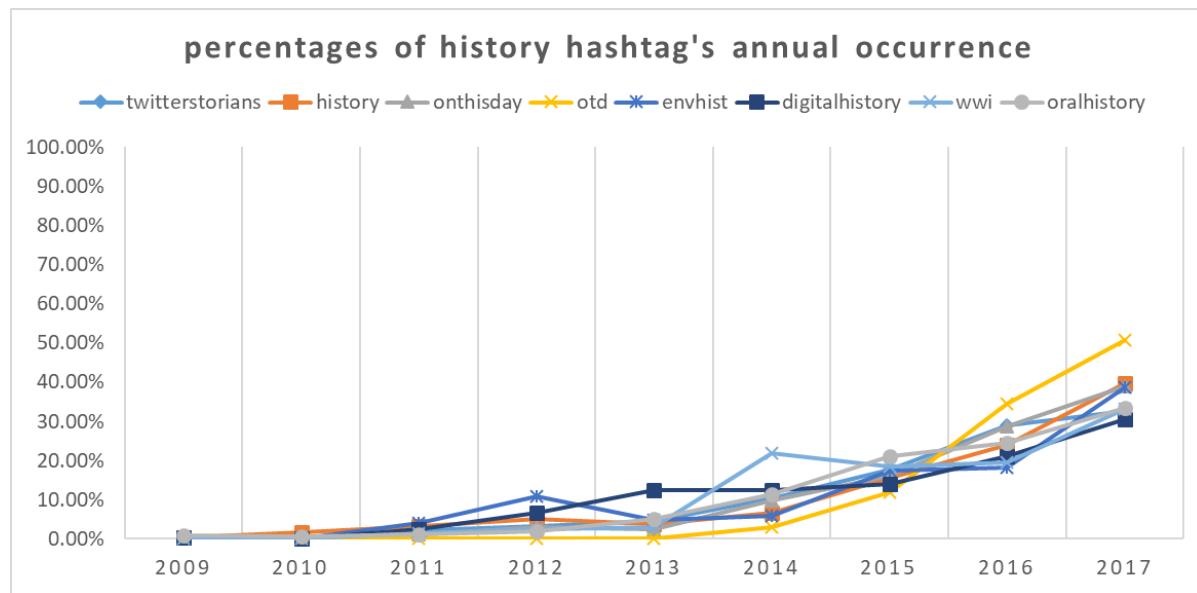


Figure 5.25: Line graph of the relative percentages of history-related hashtags' annual occurrences in the current dataset (2009-2017)

According to Figure 5.25, the interests of history-related topics have been growing gradually and steadily since 2009. The line graph does not show any obvious turning points, and it indicates a steady growth of people becoming interested in history or joining the DH Twittersphere over the past decade. It seems that the trend will continue to be a steady growth in the near future. Does this also indicate that there will be more history-related publications in DH soon? It might be too early to answer the question. However, the beginning of the rise of history-related research interests can be found in recent DH publications, and many have encouraged more history studies in DH, although there is still a long way to go (Zaagsma, 2013; Nyhan and Flinn, 2016).

Compared to the line graph of history hashtags, the line graph for education is relatively more fluctuating (see below Figure 5.26). Although the education-related hashtags have also attracted increasing attention during the period, the interest seems to decline in 2016 and 2017. Will the education-related topic lose scholarly interests in the future as may be shown in the line graph? As we know there are more and more education-related articles and books being published recently, e.g., (Lubek et al., 1995; Stutsman, 2013; Veletsianos and Kimmons, 2016), but whether there will be a downturn in the DH topics preference remains to be seen.

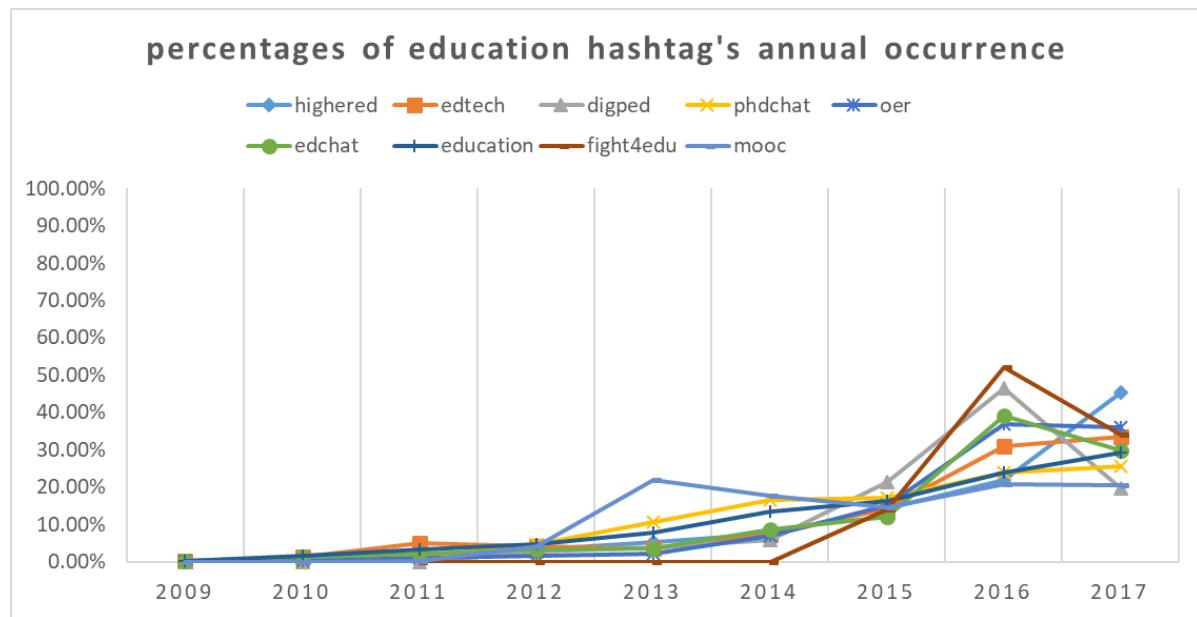


Figure 5.26: Line graph of the relative percentages of education-related hashtags' annual occurrences in the current dataset (2009-2017)

The line graph of politics-related hashtags is very different to that of history and education (see Figure 5.27).

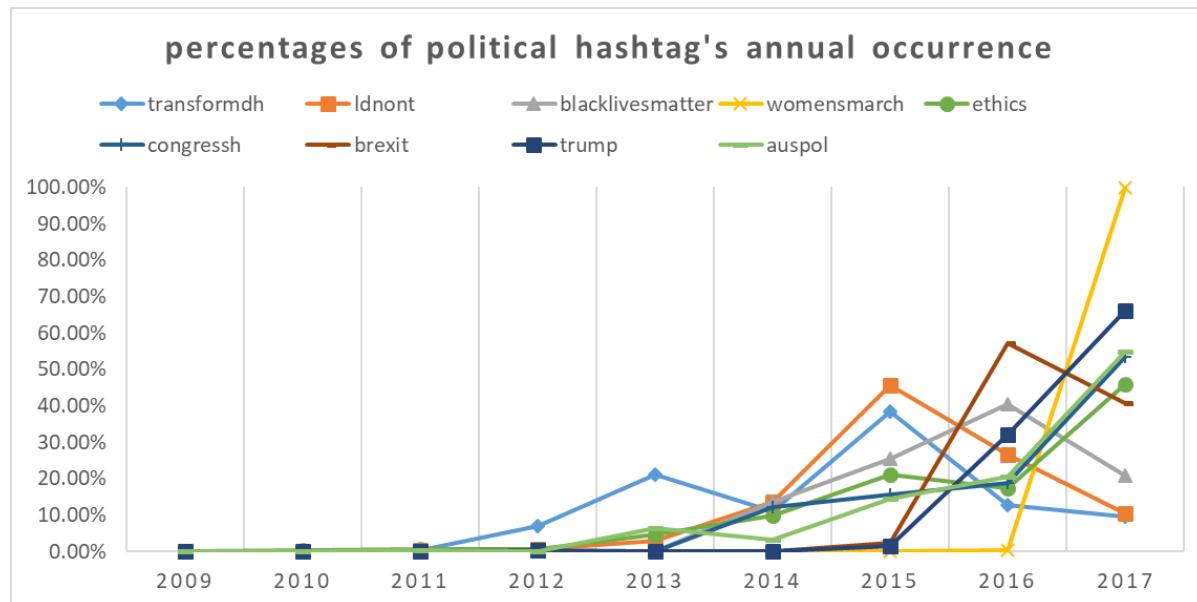


Figure 5.27: Line graph of the relative percentages of politics-related hashtags' annual occurrences in the current dataset (2009-2017)

As shown in Figure 5.27, different hashtags attract attention in different years, presumably depending on political events. Although it still shows a general increasing

trend, the development tendency is distinct from hashtag to hashtag. Apart from the reason that there are more and more general users joining Twitter, which brings more shares of interests to each topic, we can see that some topics are experiencing downturns while others are rising sharply. For example, ‘transformDH’ was firstly used in 2011 (Bailey et al., 2016), and it gradually reached its peak in 2015 before a rapid decline over the next two years. ‘Womensmarch’, on the other hand, only appeared in the year 2017 (and onwards, although the dataset does not cover beyond 2017), and it therefore, reached to 100% in 2017 in the line graph being the highest. Moreover, the network shows clear evidences of political news involved in mostly Anglophone countries, such as the US (e.g., ‘trump’, ‘congressh’), the UK (e.g., ‘brexit’), Canada (e.g., ‘ldnont’), Australia (e.g., ‘auspol’), while there are hardly any political keywords in non-Anglophone countries or regions.

Generally, unlike history and education hashtags, it is difficult to use the line graph of political hashtag usage to forecast the tendency for topics, as it depends heavily on the current breaking news (particularly in Anglophone countries). It is also difficult to predict if a growing trend of a certain hashtag will continue increasing or reach its climax in the future. On the other hand, it is difficult to see political topics occupying an obvious position in the co-citation network or groups of politics-related publications in the current dataset. Given the very active discussions on DH Twittersphere, though, there might be some to come.

Nevertheless, talking politics on Twitter is not a DH-specific character, it is common among all users as more and more campaigns and movements circulate on Twitter (McGregor and Mourão, 2016). While many DH scholars are talking about politics, most of these tweets are not related to their research, but merely personal interests. This also agrees with Grandjean’s argument that many scholars on Twitter share and discuss information that are not research-related (Grandjean, 2016, p. 2).

In conclusion, can Twitter hashtag usage help to forecast the field’s future development? It is a question to be continually investigated, and future studies can combine data from conference papers that contain work in-progress. As reviewed earlier in section 1.1.2, DH has been criticised for its many ‘ills’ including the lack of political commitment and unbalance between research and teaching (Gold, 2012, p. xii). The popular topics on Twitter, however, demonstrate something different, and by

knowing that, we might be able to see these topics appear in publications soon. Additionally, if we combine the observation made in 5.2.6.3 (Section C – Non-Anglophone DH), it seems that Anglophone DH users are more likely to discuss humanistic topics on Twitter (e.g., history, education, politics) while non-Anglophone users are more likely to discuss technical topics. As mentioned earlier, in the coming section 5.3 (Co-retweet network), this study will combine the language and affiliated country data to continue discussing the DH Twitter community.

5.3 Co-retweet network

The co-retweet network is the last of the four networks that this thesis constructs. As discussed earlier in Methodology (chapter 3), to visualise the DH scholarly social connection as well as the retweet interest, this study chooses the co-retweet link as the key connection to build the scholarly social network of the DH community on Twitter.

The network was constructed by calculating the number of non-self retweets that the 3,154 users have received (*non-self retweet count*) to weight the nodes, and the number of same tweets that any pair of users both retweeted (*co-retweet count*) for the edges. The details are explained in section 5.3.1 (Node) and 5.3.2 (Edge), respectively. Network construction procedures are demonstrated in section 5.3.3 (Network visualisation), and centrality measures are presented in section 5.2.4 (Centrality). Unlike the bibliometric networks in chapter 4 where environmental factors (e.g., affiliated country and gender) were presented in individual sections, this co-retweet network combines them in section 5.3.3 (Network visualisation) and 5.3.5 (Longitudinal and diversity analysis) to better explain the formation and development of the DH Twitter networks.

5.3.1 Node

Different users have a different contribution and influence towards the Twitter communities and other users. Twitter visualisation is one way to represent quantified DH user contribution, connection and background, such as the study done by (Grandjean, 2016).

As mentioned, among all the features on Twitter (e.g., *follow*, *mention*), *retweet* often indicates the importance and influence of the content and its level of information that is circulated to broader audiences. Therefore, *retweet* generally indicates larger impact than *like* or *mention* (Suh et al., 2010). This study has calculated the weight of user nodes by the number of non-self retweets that each user received in the current dataset.

Similar to the calculation of non-self citation in section 4.2 (ACA network), this study firstly counted the total number of retweets (including retweet and quote retweet) that a user received in the dataset. Then, the number of self retweets (i.e., the number of times this user retweeted themselves) was removed from the total number. As can be seen from Figure 5.28 and Figure 5.29, the number of non-self retweets on Twitter has a power-law-like characteristic (Clauset et al., 2009) – a few users received extensive retweets whereas most users were not retweeted or were only retweeted a few times. This power-law distribution is also known as one of the prerequisites of a ‘small-world’ type of network (Bork et al., 2004).

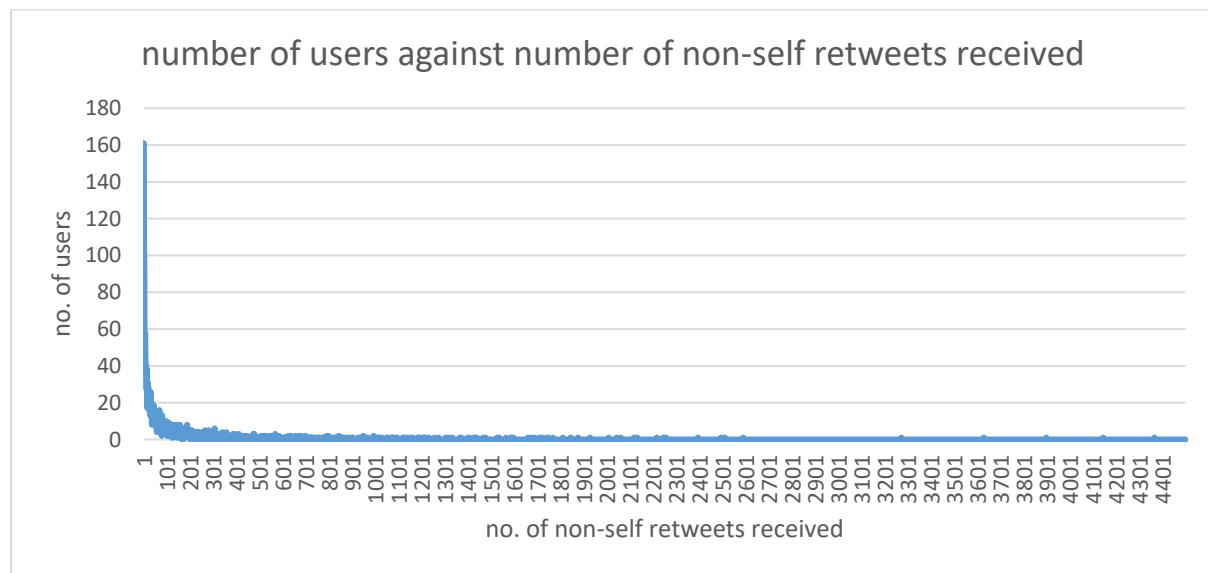


Figure 5.28: Line graph of number of users against number of non-self retweets received

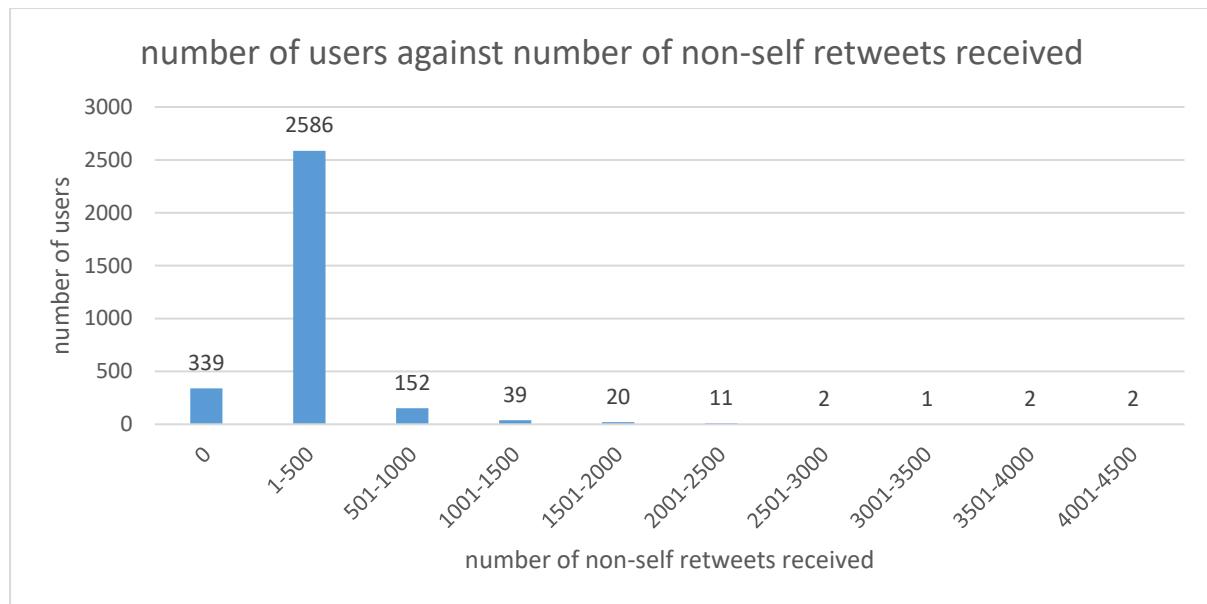


Figure 5.29: Bar chart of number of users against number range of non-self retweets received

As the data in Figure 5.28 and Figure 5.29 shown, only 77 users (2.44%) were retweeted more than 1,000 times (non-self), while 339 users (10.75%) did not receive any non-self retweets and 2,586 users (82.00%) only received less than or equal to 500 non-self retweets. This power-law-like result agrees with the findings of Ediger et al. that retweets, in general, tended to come from a relatively small group of original tweets (Ediger et al., 2010). The DH Twitter study done by Ross et al., has a similar argument, seconding the Nielsen ‘90:9:1’ rule, i.e., 90% of users are ‘lurkers’, 9% of users contribute from time to time, and 1% participate a lot and account for the majority of contributions (Ross et al., 2011, p. 221). Also, some explained this as ‘the more you tweet, the more retweets you will usually get’ (Bullas, 2013).

In total, this study counted the non-self retweets of 3,154 DH users, and Table 5.6 below shows the top 20 users who received the most non-self retweets during all years (2009-2017) as well as in individual years. This study has also collected the data for 2006-2008, however, there are no retweets in the dataset and so the numbers for users are all zero.

Table 5.6: The top 20 users who received the most non-self retweets during all years (2009-2017) and in individual years.

No.	User name	All years (2009-2017)	2017	2016	2015	2014	2013	2012	2011	2010	2009
1	Bethany Nowviskie	4,366	1414	699	596	511	476	348	223	92	7
2	NEH	4,145	1708	888	694	437	178	142	53	45	0
3	melissa terras	3,899	1010	680	553	591	562	294	144	61	4
4	Dan Cohen	3,629	555	647	572	490	520	370	303	160	12
5	Miriam Posner	3,273	951	691	630	518	195	189	98	1	0
6	ADHO	2,590	786	989	530	198	87	0	0	0	0
7	DigitalHumanitiesNow	2,510	598	321	482	386	311	293	90	28	1
8	The DLF	2,494	949	697	513	231	78	17	9	0	0
9	Alex Gil	2,395	782	750	390	272	149	34	17	1	0
10	Tim Sherratt	2,257	886	834	287	79	41	59	54	16	1
11	HASTAC	2,248	527	377	326	271	256	140	141	209	1
12	Brian Croxall	2,244	687	446	294	323	208	165	88	32	1
13	Mia R	2,217	606	553	387	324	210	76	54	7	0
14	Ted Underwood	2,131	846	494	347	228	130	77	9	0	0
15	Trevor Owens 	2,118	772	471	371	277	87	87	48	5	0
16	Marin Dacos	2,061	640	486	188	142	194	152	116	136	7
17	Matthew Kirschenbaum	2,043	618	425	203	291	223	111	128	42	2
18	Jesse Stommel	2,008	386	729	468	279	123	20	3	0	0
19	The Chronicle	1,928	554	486	350	190	102	86	99	58	3
20	State Library of NSW	1,877	894	523	245	132	60	23	0	0	0

As shown in Table 5.6, to represent individual users, this study chooses to display the username instead of the Twitter handle for better recognition and comparison with previous networks (e.g., bibliometric networks done by this study in the previous sections). By looking at the number of non-self retweets, we can see that the number varies even among the top 20 users. While some users joined Twitter as early as in 2009 (e.g., Bethany Nowviskie, melissa terras and Dan Cohen), others started to receive retweets only after 2013 (e.g., ADHO).

5.3.2 Edge

To calculate the edge, this study has counted the occurrences when two users co-retweeted the same tweet. For example, user A retweeted a tweet, and user B also

retweeted the same tweet, then their co-retweet value increases by 1. The co-retweet value is the number of same tweets that they both retweeted.

It should be noted that it is the same tweet that both users retweeted, not the same user (or author). Some co-retweet studies (e.g., Wang et al., 2014, p. 9), counted the number of users (instead of tweets) that have been co-retweeted. Wang's method counted the occurrences when user A retweeted a tweet that was originally posted by user X, and user B retweeted another tweet that was also originally posted by user X but that was different to the tweet user A retweeted, and thus Wang counted that the co-retweet value of user A and B increased by 1. This method employed by Wang is questionable, and it, potentially, will not only fail to visualise the topical relevance on Twitter, but will also provide confusing mapping that might misrepresent the Twitter community. As presented in the hashtag network (section 5.2, Hashtag co-occurrence network), the same user might tweet on various topics, e.g., research, politics, news. Retweeting the same author of different tweets ignores the topical diversity a user on Twitter might have, and such a method should be avoided when constructing a co-retweet network. Moreover, Wang et al. themselves also acknowledged that each tweet should be treated as a single document which contains similar topics, and that not all the tweets of a user should have similar topics when they calculated the tweet topic by LDA, i.e., the Latent Dirichlet Allocation (Wang et al., 2014, p. 11).

5.3.3 Network visualisation

By using the VOSviewer, this study visualised the co-retweet network of 3,154 DH Twitter users in the current dataset (see Figure 5.30). In total, 2,982 users (94.55%) are connected to the network, and similar to the hashtag network, this percentage is significantly higher than previous bibliometric networks.

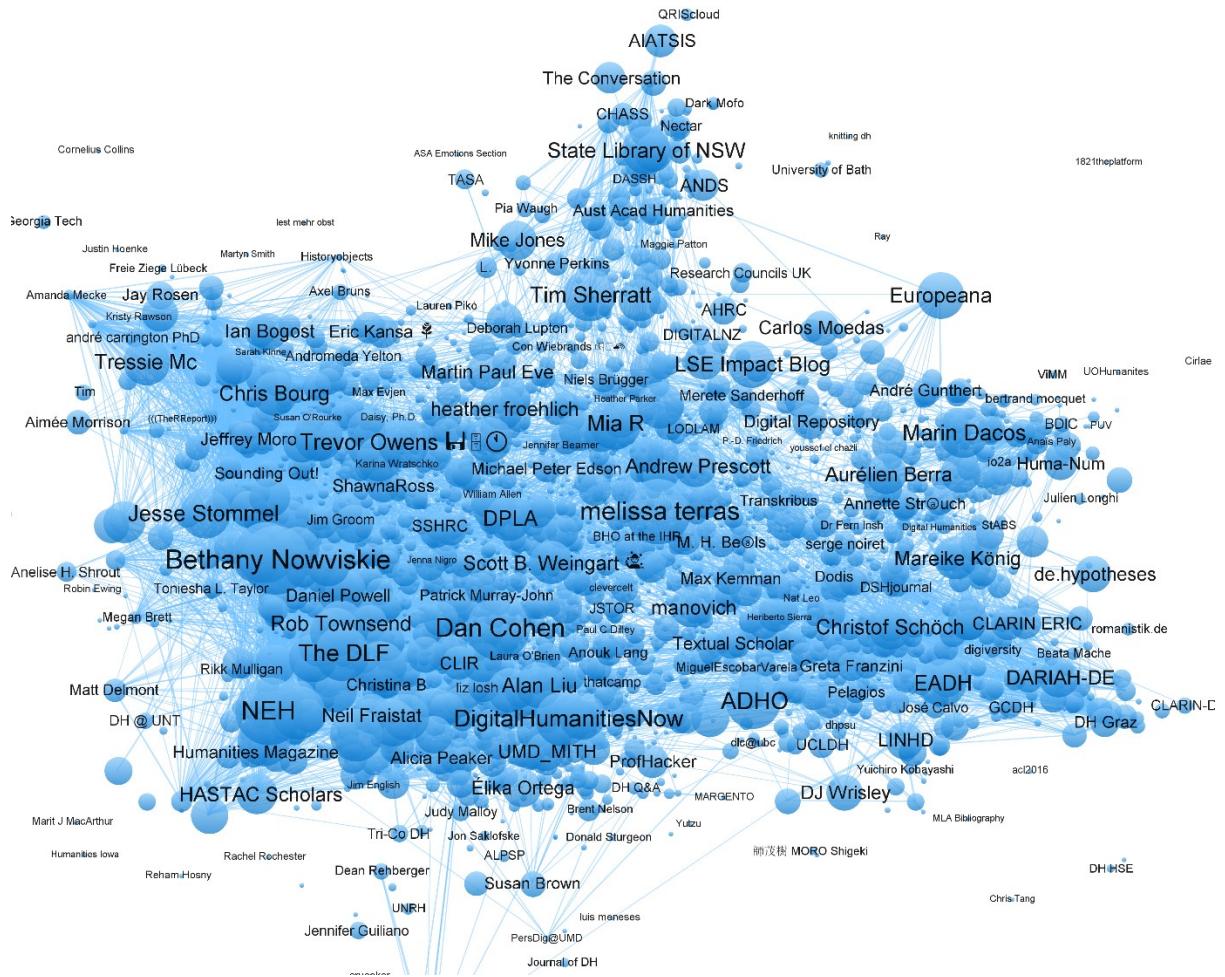


Figure 5.30: User co-retweet network in DH, data extracted from Twitter.

As can be seen in Figure 5.30, the network, in general, is relatively dense compared to the bibliometric networks. There are several apparent clusters spread across the map, and the biggest and most dense cluster on the left includes many users based in North America, such as Bethany Nowviskie, NEH, Jesse Stommel. In the middle, there is a cluster with many UK based users, such as Melissa Terras, Mia R, and Andrew Prescott. On the bottom right corner, there are many organisational accounts (e.g., ADHO, EADH, Dariah-DE) with users mostly from German-speaking (e.g., Christof Schöch), Dutch-speaking (e.g., Max Kemman) and Spanish-speaking regions (e.g., LINHD). Above this cluster on the top right corner, we can see users from French-speaking regions, such as Marin Dacos and Huma-Num⁶⁴. At the very top of

⁶⁴ Huma-Num, a large research infrastructure for SHS (the Humanities and Social Sciences) in France. More information can be found: <https://www.huma-num.fr/>

the network, there is a cluster with many users based in Australia, for example, Tim Sherratt, State Library of NSW, and AIATSIS⁶⁵.

The node distribution in this network mainly agrees with the ‘small world’ finding in Grandjean’s DH user following network (Grandjean, 2016, p. 4), with only a few DH users receiving an extensive amount of retweets whereas most users were not retweeted or were only retweeted a few times. Thus, this study revisits the question that was asked by Grandjean: ‘are digital humanities – whereas describing themselves as a transversal field – finally a closed world where everybody knows everybody?’ (Grandjean, 2016, p. 4).

Apparently, the co-retweet network (in Figure 5.30) is indeed a small-world network and has many characteristics of such a type of network. However, it is no longer as dense as the network constructed by Grandjean in 2014 which had only two clusters (one large dense cluster and one small cluster). Therefore, it may still be a small world where many people know everyone, but the network has clearly been developing since 2014. From the current network, we can see evidently that there are several connected clusters with obvious boundaries. Users in the network might still know most users, but their retweeting behaviours are in favour of a certain group of tweets instead of all the contents posted by most of these users. It no longer seems to be a centralised map where all users are clustered together.

Additionally, in Grandjean’s study, these users were collected based on the ‘follow’ decisions made by prestigious DH organisational accounts, and such decisions could be selective and act as a filter that includes mostly the ‘insiders’ and ‘big names’ that might already know each other very well. Newcomers and lesser-known practitioners, on the other hand, might be excluded in such a ‘follow’ list and thus were excluded in Grandjean’s dataset. The current data collection potentially contributes to the small-world argument and makes the case even stronger.

Moreover, as mentioned earlier, Ediger et al. found that retweets, in general, tended to come from a relatively small group of original tweets, which means that the small-

⁶⁵ AIATSIS, Australian Institute of Aboriginal and Torres Strait Islander Studies. More information can be found: <https://aiatsis.gov.au/>

world networks on Twitter are typically very common (Ediger et al., 2010), and the DH community is not an exception. This will be further discussed at the end of this section (5.3.6).

5.3.4 Centrality

To better investigate the roles that scholars play in the Twitter community, centrality measure can be beneficial. Similar to the previous networks, Gephi 0.9.2 has been employed to calculate the *betweenness centrality*, and the complete table of results can be found in Appendix E.

5.3.5 Longitudinal and diversity analysis

User interest is believed to be dynamic and changing over time on Twitter (Ahmed et al., 2011; Yin et al., 2011). This study has, thus, visualised the structure of the Twitter network in each individual year longitudinally (2009-2017). Similar to the Hashtag networks, there is no network that can be constructed before 2009. This might be because it is difficult to trace the early practice of retweet usage (i.e., adding 'RT' in front of the original tweet). As an important part of DH history, the DH user activities on Twitter are complementing and influencing its development, and by adding diversity information (i.e., affiliated country and gender) to the individual years, we can observe and examine how the community had been formed on this social media platform.

5.3.6 Discussion and analysis

This section discusses the social network and co-retweet pattern of DH scholars on Twitter. Similar to previous networks, it does not merely focus on identifying the scholarly groups, but more importantly, their connections and the structure of distribution. It addresses the *scholar* and *environment* research questions from three perspectives – 5.3.6.1 Country, 5.3.6.2 Historical periods, 5.3.6.3 Gender, and each considered a factor that influences the formation of the DH co-retweet network. The three sections are discussed in a different order to that of the co-authorship network due to the different structures of the two networks.

By reviewing the results, we can see that the DH social network based on retweet activities is formed mainly by user interface languages and locations, and that not all regions joined the network at the same time. Users in Anglophone countries are the

majority in the dataset, and holding international DH events encourages local users as well as users in other countries to join in the Twittersphere, especially when the organiser is in a non-Anglophone country. In terms of gender, although male and female scholars account for the same proportion, most central positions are taken by female scholars who have been acting as critical bridges in forming connections.

It needs to be noted that people retweet for a variety of different reasons and this study only focuses on two factors (country/language and gender). There are many other important factors, such as for commenting, validating, socialising, gaining visibility, etc., (Quan-Haase et al., 2015a, p. 3).

5.3.6.1 Country

In order to dig into the structure of the co-retweet network more, this study has colour-coded the nodes according to users' interface language (see Figure 5.31).

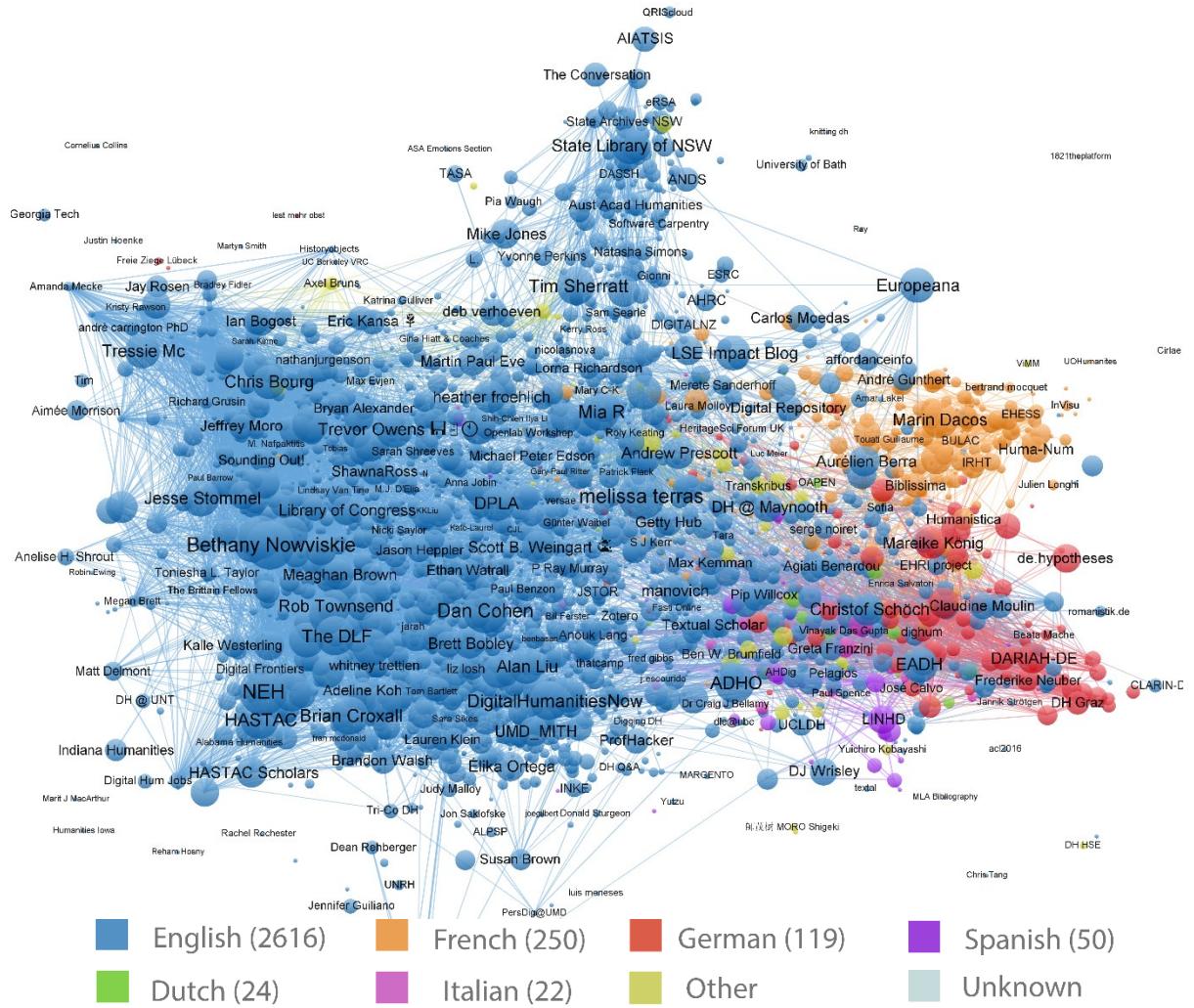


Figure 5.31: User co-retweet network in DH colour-coded with different languages, data extracted from Twitter.

As we can see from Figure 5.31, the formation of the co-retweet network is highly correlated with language use, especially for English, French, German and Spanish speakers. The majority of users (82.94%) are using English (coded in blue) as their Twitter interface language, while 7.93% and 3.77% users are using French (in yellow) and German (in red), respectively. It seems on the network that users using the same language tend to retweet similar tweets, although it is no surprise as the majority of tweets are text-based information.

Such clustering is more apparent when the nodes are colour coded with affiliated country information (see Figure 5.24).

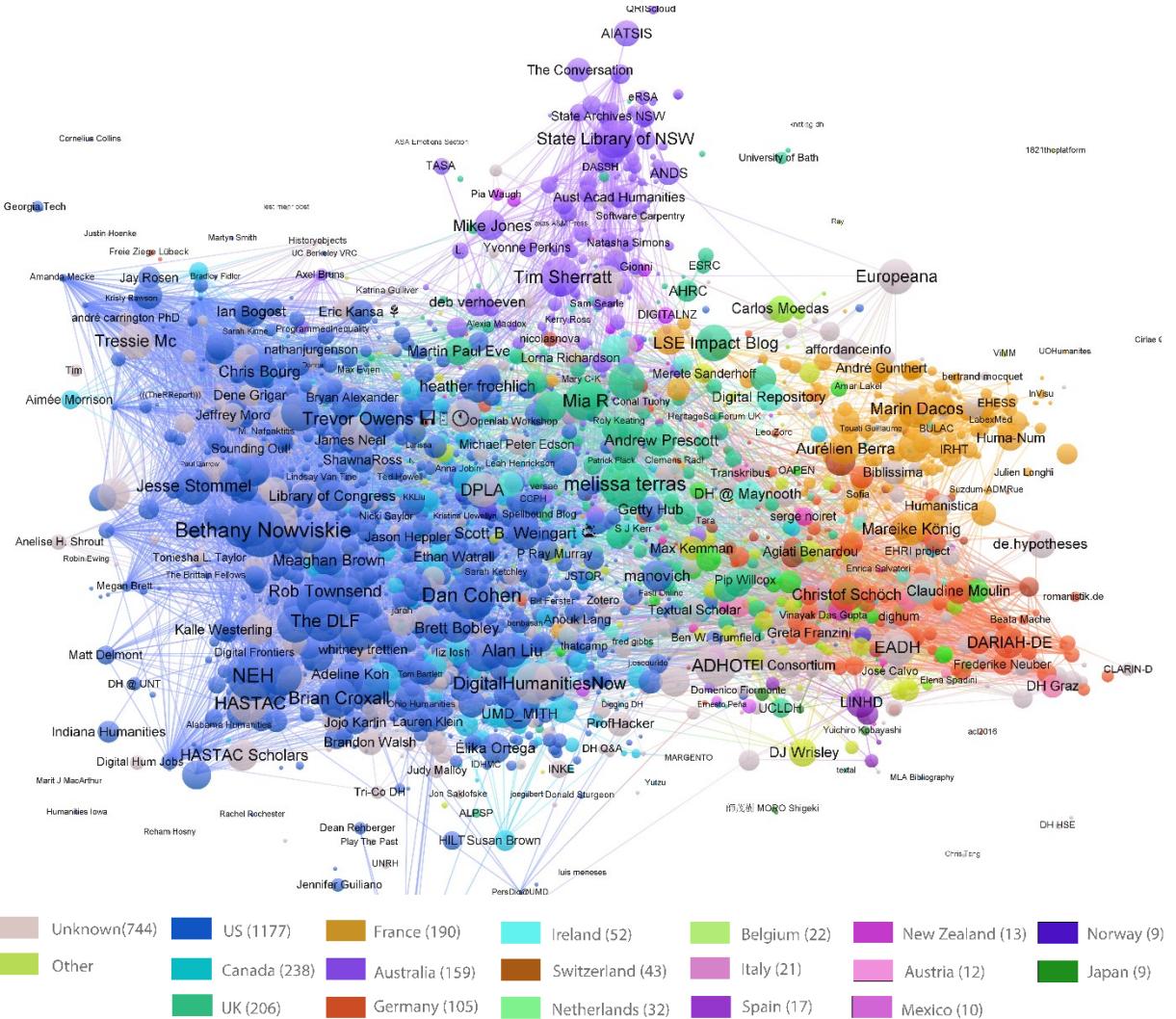


Figure 5.32: User co-retweet network in DH colour-coded with different affiliated countries, data extracted from Twitter.

Figure 5.32 shows a detailed node distribution, and with the added country information, one can distinguish the grouping within the English-speaking regions. Apart from the other European language-speaking countries, users from the USA (65.64%) account for the most part of English-speaking regions on the left of the network (in dark blue), the UK cluster (in green) is in the middle, the Canada nodes (in light blue) are scattered around the USA and UK clusters, and the Australian and New Zealand clusters (in dark and light purple) are at the top of the network. Among these five English-speaking countries, the users from UK and Canada are more spread across the network and mingled with other clusters, while users based in the USA and Australia are more likely to retweet the same contents with their local peers.

As users from the USA and Canada account for a large cluster on the left, it is worth investigating the retweeting patterns across different time zones in these two countries. Below is the network that is mainly colour-coded with time zones in the USA and Canada (Figure 5.33).

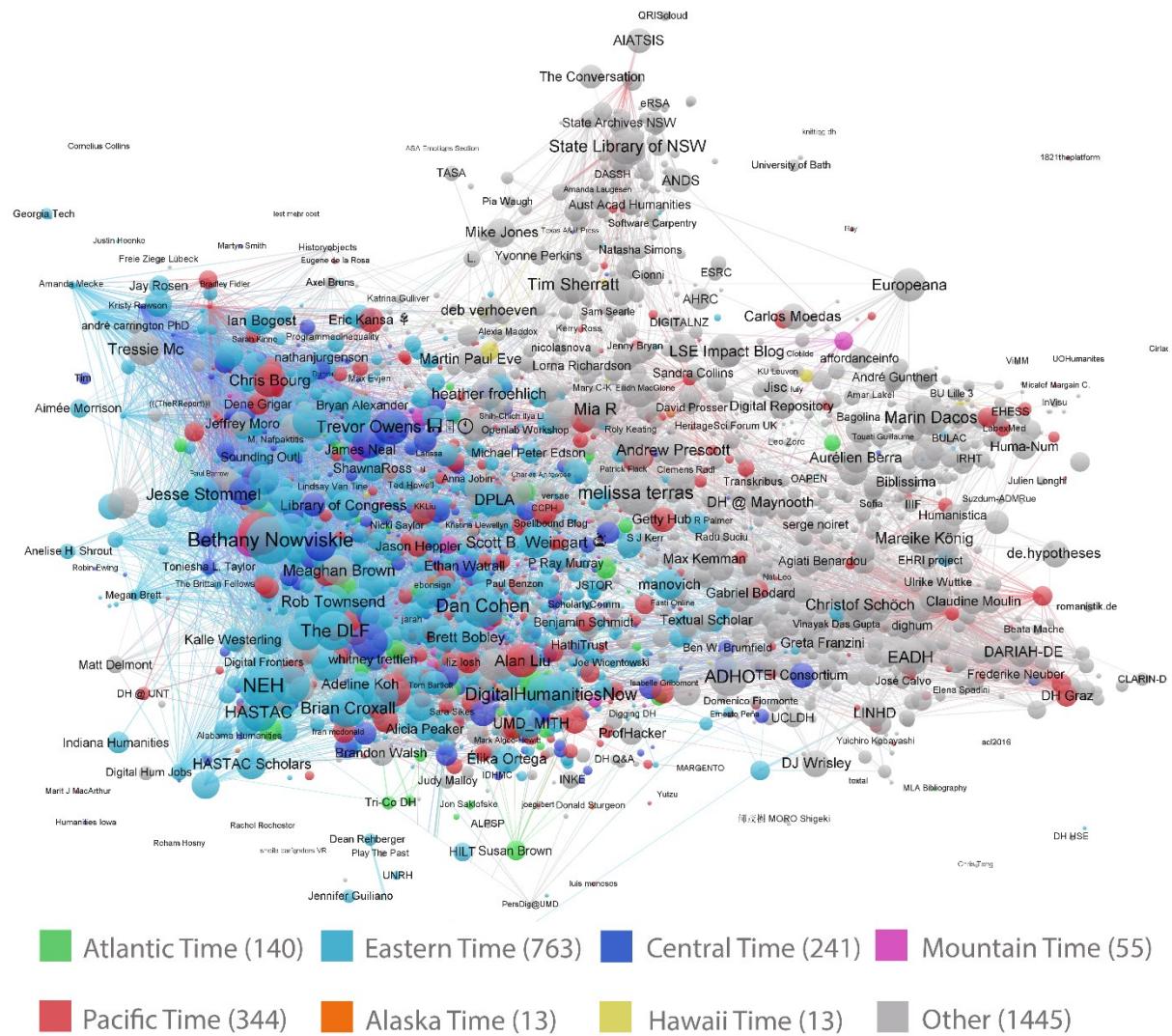


Figure 5.33: User co-retweet network in DH colour-coded with different time zones in the USA and Canada, data extracted from Twitter.

As seen in Figure 5.33, among all the users that are in the above time zones, people from the East of North America are more clustered on the left of the network (within the USA cluster) while people from the West of North America are more positioned across the whole network. Most users in the USA and Canada are from the East. For example, there are 763 users (48.63%) in the Eastern Time Zone (in light blue), 140 users (8.92%) in the Atlantic Time Zone (in green), and 241 users (15.36%) in the

Central Time Zone (in dark blue). These users in the East are mainly positioned within the USA cluster on the left of the network, which shows that they have similar retweet interests as their local peers in North America. However, although users in the West are in regions with a smaller population, they spread more across the whole network, which means they have broader retweet interests. For example, the second largest user group is in the Pacific Time Zone (in red, 21.92%), and unlike the Eastern Time Zone group, many of them can be found in the UK, French and German clusters (in red), although many are still focused on the left part. The same with other time zone users in the West of North America, such as Mountain Time, Alaska Time, and Hawaii Time.

This is an interesting phenomenon that might raise new questions for future social media studies in the North America. This study, however, chooses not to engage further because it only explores regional differences of DH user groups on Twitter.

Back to Figure 5.31 and Figure 5.32, it is clear in the network that users from the same location and/or using the same language tend to cluster. Anglophone countries make up the majority (82.94%) of this network. On the top right, there is a cluster of French-speaking users (in yellow) accounting for 7.93% of the total users, while on the bottom right is a cluster of German-speaking users (in red) making up 3.77% of the total users. There are also a few users from other language groups, e.g., Spanish (50 users, 1.67%), Dutch (24 users, 0.80%), Italian (22 users, 0.74%).

Such country distribution matches many previous DH studies (e.g., Dacos, 2012, 2013; Grandjean, 2016; Weingart and Eichmann-Kalwara, 2017). For example, Dacos has presented a 'Digital Humanities Decision Power' indicator that measures a country by their number scholars in positions of power and decision-making (e.g., as reviewers) and the results show that the UK, Ireland, USA, Canada, and Australia were the five leading countries, and all are Anglophone (Dacos, 2012, 2013).

To understand what DH users are retweeting about, Table 5.7 below shows an example of the 20 most retweeted tweets in the current dataset (for the complete table of 200 most retweet tweets, see Appendix E). The number of retweet counts shown in the table are the number of times the tweet has been retweeted by the current

collected group of DH users, not the total number of times this tweet has been retweeted by all users on Twitter.

Table 5.7: The 20 most retweeted tweet by selected users in the current dataset

	Retweet count	Username	Given location	Tweet
1	91	Matt Gold	United States	Please help us spread the word!! New CFP for Global Debates in the Digital Humanities #DH2017 #dhdebates https://t.co/cq1nLhNguH
2	62	amardeep singh	United States	DH friends: Lehigh is hosting a digital humanities/social justice themed conference in April 2018. Spread the word! https://t.co/2EGqVFiY2c
3	55	Matt Gold	United States	.@laurenfklein & I are excited to release the open interactive ed of Debates in the Digital Humanities 2016! https://t.co/HutjiVN2II #dh2016
4	54	DH2017 Montreal	Unknown	CFP for #DH2017 in Montreal now live! https://t.co/b2QTcRfOFR Appel à communication #DH2017 à Montréal disponible ! https://t.co/0a45WTfd9j
5	53	NEH	United States	NEH announces \$39.3 million for 245 humanities projects and programs. #NEHgrant https://t.co/b4siNFr3QQ https://t.co/VrjTSWJQHg
6	52	Jason Scott	Unknown	Spread the word: The Internet Archive has put up *25,000* 78rpm records, digitized professionally. https://t.co/y3kUOcdIWZ
7	51	NEH	United States	Now available: NEH emergency grants of up to \$30,000 for cultural institutions impacted by Hurricanes Harvey & Irma https://t.co/rWpVpxqawU
8	51	Internet Archive	United States	If you see something, save something. Use Save Page Now to save URLs at https://t.co/qGDMcsIRtF https://t.co/AtHn0UhREo
9	50	Julianne Nyhan	Unknown	New book by @Andyucl & me on #oralhistory of #digitalhumanities is published & available open access! Get it from: https://t.co/RJaIGyY7Qe
10	50	Élika Ortega	United States	Delighted to share the #dh2018 CfP in Sp, En, Fr & Pt. Italian & German coming soon! Please share widely! https://t.co/rwjE7mHMWK
11	49	Zotero	Unknown	Introducing Zotero 5.0: My Publications, Feeds, improved syncing, improved browser connectors, and much more! https://t.co/M9yG5Dnfui
12	47	Miriam Posner	United States	.@alison_booth and I are coordinating a special issue of PMLA on "varieties of DH." Papers due 3/18. https://t.co/jJkJcKikhq #DH2017
13	47	Manifold Scholarship	United States	We are delighted to launch the beta version of @ManifoldScholar, a new open-source platform for scholarly publishing https://t.co/TzWiTqnfZL

14	46	Ted Underwood	United States	In this article, I'm accused of not listing "interesting things"--something I can rectify in a Twitter thread! 1/? https://t.co/zPPuRVjaeT
15	46	Rob Townsend	United States	"#NEH will not invite or accept any new grant applications, nor make any new grants, in FY 2018" https://t.co/30H6anHL13
16	45	dh+lib	Unknown	"Open Stacks: Making DH Labor Visible" - a new piece by @laurabrarian https://t.co/kgHtMqkLF1 #digitalhumanities #libraries
17	45	Stewart Varner	United States	.@pmhswe & I wrote for @amlibraries but there were some issues w/ the editing. Here's what happened: https://t.co/y85Dbhjly7
18	44	Christof Schöch	Germany	#DH2017: individual #abstracts (PDF) and full book of abstracts now available here! https://t.co/24ZgfWNhyN
19	43	dhdebates	Unknown	New CFP: Global Debates in the Digital Humanities, Eds @fiormont @paolaricaurte & Sukanta Chaudhuri https://t.co/2oqAqxYdcQ #dhdebates
20	42	Alex Gil	United States	Black Digital Humanities Projects & Resources at 100+ resources now. Just added ~20 myself. Incl. paywalled https://t.co/JBcqG6pmGQ

From Table 5.7, we can see that all the tweets are written in English, even the ones posted by the user in Germany or the users that do not give their location information on Twitter. 13 out of the 20 tweets are posted by users located in the US. Another interesting finding is that all these tweets include links, and some of them are quote retweets. However, are these features helpful for getting more retweets? It is still too early to conclude that tweets that are written in English, include links, and are posted by users located in the US are more likely to attract retweets from DH users.

As this study has noted earlier, the proportions of each country presented on this network do not equal the actual DH communities on Twitter. Users were selected according to the 'follow' list of ADHO (and its members) Twitter account. Not only is English the language that these organisations use most of the time on Twitter, many Anglophone countries are believed to hold the largest number of memberships and leadership positions at ADHO (Fioronte, 2012, p. 6). Because of such a data sample, this analysis does not necessarily indicate that the Anglophone community is the most active DH group.

The lack of non-Anglophone representation is an issue that has been raised by many scholars appealing for more diversity in DH, as mentioned earlier (e.g., Fioronte,

2014; Galina, 2014; Mahony, 2018). New measures are needed to further investigate the network, and the calculation of average node weight and betweenness centrality can help to level such representation (see below Table 5.8).

Table 5.8: 10 countries with most users ranked by the value of average *betweenness centrality* on the co-retweet network. Table also includes average node weight and average number of tweets, data extracted from Twitter, 2006-2017

	country	no. user	average betweenness centrality	average weight	average no. tweets
1	Australia	158	2,752.824574	174.360759	2,319.101266
2	United Kingdom	206	2,493.249441	164.097087	2,071.519417
3	France	190	1,705.762141	140.473684	2,187.089474
4	United States	1,176	1,666.148717	163.984694	2,036.976190
5	Netherlands	32	1,640.628288	155.906250	1,402.843750
6	Germany	105	1,379.571603	152.133333	1,720.914286
7	Canada	237	1,262.095294	97.514768	1,726.713080
8	Ireland	52	990.188001	137.711538	1,955.134615
9	Belgium	22	592.239121	111.318182	1,815.863636
10	Switzerland	42	301.876359	108.880952	1,322.428571

Table 5.8 is ranked by the average betweenness centrality, and it shows the top 10 countries that have the most users in the current dataset. It also includes the average node weight (i.e., number of non-self retweets received) and the average number of tweets that users in each country have produced. Although the sample size of each country varies (e.g., the US has 1,176 users while Belgium only has 22 users) and the limitation is noted, average values do provide a new aspect to see central positions and node importance on the network.

The first interesting thing to see on Table 5.8 is the great influence and central position of users located in Australia and the UK. Despite that Australia and the UK only account for 6.90% and 5.30% of the total users (which are far less than the US, 39.44%), they have the highest average weight and betweenness centrality. As discussed in chapter 5 (5.3 Co-retweet network), node weight is calculated based on the number of non-self retweets the user has received, i.e., the higher the weight, the more times the user has been retweeted by the users in the current dataset, indicating

the user's influence and contribution to the DH Twitter community. The node *betweenness centrality* counts the number of the shortest possible paths of any other two nodes passing (or 'between') a node (Badar et al., 2013, p. 759), i.e., the higher *the betweenness centrality*, the more important and central the position of this node on the network. In the co-retweet context, a user who has higher *betweenness centrality* is often retweeted by a wide range of users from different clusters so that the user can remain in the central location being the bridge to connect different clusters.

Looking at the averages, why do so many people from different countries and language groups all retweet the tweets posted by Australian and UK scholars? Admittedly, English is a language that seems to bring advantages to such retweet preferences, but why not the US (which has the most users) or Canada (which has the second most users)?

If we look at the last column on Table 5.8 – the average number of tweets per user, we can see that users in Australia and UK have the highest and the third highest numbers of tweets (France ranks as the second highest, but it is not an Anglophone country, and it will be discussed separately). It is said that the number of tweets a user posts often correlates with the number of retweets they get, and some suggest that 'the more you tweet, the more retweets you will usually get' (Bullas, 2013; Jenders et al., 2013). To verify such a statement statistically, this study has calculated the *Pearson correlation coefficient* between the number of tweets a user posted and the number of non-self retweets the user received (i.e., the node weight) of all users (3,154 users) in the current dataset. The value is around 0.3205, which indicates a moderate positive correlation (Taylor, 1990)⁶⁶, and the complete table can be found in Appendix E.

As the *correlation coefficient* value shows, although the number of tweets do not strongly (or completely) link to the number of retweets, demonstrating only a moderate

⁶⁶ According to studies (Bartko, 1966; Taylor, 1990), when the correlation coefficient value r equals to: a. 0, then no linear relationship; b. +1 indicates a perfect positive linear relationship; c. -1 indicates a perfect negative linear relationship; d. Values between 0 and 0.3 (0 and -0.3) indicate a weak positive (negative) linear relationship; e. Values between 0.3 and 0.7 (-0.3 and -0.7) indicate a moderate positive (negative) linear relationship; f. Values between 0.7 and 1.0 (-0.7 and -1.0) indicate a strong positive (negative) linear relationship.

positive correlation, it is not difficult to understand why Australian and UK scholars have got the most retweets – they have tweeted more often than users in other countries. In the current dataset, on average, each Australian user has posted 2,319 tweets (ranking 1st) and each UK user has posted 2,072 tweets (ranking 3rd); on average, each Australian user has received 174 non-self retweets (ranking 1st) and each UK user has received 164 non-self retweets (ranking 2nd). The ‘moderate’ (instead of ‘strong’) degree of *correlation coefficient* (0.3205) also explains some level of deviation between the two variables, signalling that there are other factors that influence the changes between the number of tweets and retweets. For example, among Anglophone countries, Ireland ranks the 5th and Canada ranks the 7th by the number of tweets, but they rank the 7th and 10th by the number of non-self retweets received, respectively. Such a drop in the rankings indicates that they are potentially losing retweets and influence compared to other countries, if normalised to the same scale.

On the other hand, the correlation between the number of tweets and the value of betweenness centrality seems to be positive but weak. The *Pearson correlation coefficient* between the two variables is around 0.1323, which shows a weak positive correlation (Taylor, 1990). Despite that the average number of tweets per user in Australia and the UK (ranking the highest and 3rd highest) indeed correlates with the value of betweenness centrality (the highest and second highest, respectively), such correlation is not shown in some other countries. Some fall between the two columns (e.g., Ireland ranks 5th by the number of tweets but 8th by its betweenness centrality) while some rise (e.g., the Netherlands ranks 9th by the number of tweets but 5th by its betweenness centrality). What do users in these countries prefer to post that could cause such inconsistency in different rankings? Why do some countries post a relatively large number of tweets but do not get corresponding central positions on the network (e.g., the US)?

As mentioned, these two values – the number of non-self retweets (i.e., the node weight) and the value of betweenness centrality – are different indicators implying distinct features of the network nodes. The former indicates how loudly the user can be heard (i.e., the user’s influence and impact on the DH Twitter community), while the latter indicates how far the user can be heard (i.e., audiences from a wide range

of clusters share the user's tweets so that the user can remain in the central location being a bridge and connecting different clusters). Some countries are not heard very loudly but reach very far. For example, Canada has the least non-self retweets among the 10 countries, but it has a relatively high betweenness centrality. France has an average node weight (non-self retweets) ranking 6th, but its betweenness centrality ranks 3rd. Some countries are heard very loudly but their voices are not heard very far. As discussed, the majority of the most retweeted tweets by DH users are tweeted by the US users, and this country ranks 3rd by its number of non-self retweets. Yet, the US users do not occupy very central positions on the network, with most of them clustered on the left of the graph, ranking 4th by betweenness centrality. Some countries are heard very loudly and very far, such as Australia and the UK, as mentioned earlier.

We know that the idea of 'the more you tweet, the more retweets you will usually get' (Bullas, 2013; Jenders et al., 2013) can partly explain why Australia and the UK hold both high numbers of non-self retweets and high values of betweenness centrality. However, it is important to investigate other factors, and more importantly, what makes the tweets by users in France (and Australia and the UK) shared more widely than tweets by the US? What kind of topics do users in each country like to tweet?

If we take a look at the highly retweeted users in the current dataset, we can see that users in the US are more likely to use hashtags that are related to local events while users in Australia, the UK, and France are more likely to use hashtags that are related to global topics (for the complete table of mostly used hashtags of each user, please see Appendix E).

For example, the most retweeted DH user in the US is Bethany Nowviskie, and the hashtag she has used the most is 'DLFforum' (233 times), which is a forum held by the Digital Library Federation (DLF) in the US (DLF, 2020). Jacqueline Wernimont is another most retweeted Twitter user in the US, and her most used hashtag is 'DF17UNT' (58 times), which is the largest DH conference in Texas, hosted at the University of North Texas in 2017 (UNT, 2017). Another highly retweeted US user, Matt Gold's most used hashtag is 'digitalGC' (42 times), and it is a movement led by the Graduate Center of the City University of New York (@GC_CUNY).

In the Australian DH Twittersphere, the most used hashtags are more globally oriented. The most retweeted Australian user, Ingrid Mason's mostly used hashtag is 'eResearch' (57 times), and Annelie de Villiers, another highly retweeted user, has used 'Archives' (80 times) the most. The user 'asa letourneau' has used 'LODLAM' (171 times, 'Linked Open Data in Libraries, Archives, and Museums') the most, and the user Alexia Maddox has used 'SMSociety' the most (139 times, 'Social Media Society').

Similarly, in the UK DH Twittersphere, the most retweeted user 'melissa terras' has used 'digitalhumanities' the most. For other highly retweeted UK users, James Baker has used 'librarycarpentry' the most, a global community teaching software (Library Carpentry, 2020), and Simon Tanner has used 'sharecarex' the most, an international conference focused on collaboration and sharing in the cultural heritage sector (Sharing is Caring, 2019).

As the highest ranking non-Anglophone country, the topics that French users have used the most are more about copyright, data standards and methods. This matches the discussion in section 5.2.6.3 (Section C – Non-Anglophone DH) suggesting a strong connection between non-Anglophone communities and the technical aspect of DH topics. The French user who received the most non-self retweets is Marin Dacos, who has used 'openaccess' the most, and Pierre Mounier, too, has used the same hashtag most often. Jean-Christophe used 'iLoveOA' the most, Régis Robineau used 'IIIF' the most, and Casilli used 'digitalLabor' the most. French users seem to be more interested in talking about data (standard and methods) on Twitter. In addition, as mentioned, technical languages in scholarly communication are usually more formalised which makes it more straightforward for non-Anglophone speakers to reach Anglophone audiences (Ammon, 2006, p. 4).

We can see that different countries have different preferences for topics that they tend to tweet. Some are more local and limited while others are more global and diverse, and their preferences seem to influence their position on the co-retweet network, whether central or peripheral. How much correlation is there between the hashtag diversity and betweenness centrality? The Table 5.9 below shows the top 10 countries by average number of unique hashtags and their betweenness centrality. The *Pearson*

correlation coefficient indicates a weak but positive correlation between the two variables ($r = 0.216365$)

Table 5.9: 10 countries with most users ranked by the value of average *betweenness centrality* on the co-retweet network. Table also includes average number of unique hashtags used, data extracted from Twitter, 2006-2017

	country	average no. unique hashtags used	average betweenness centrality	no. users
1	Australia	430.7152	2752.824574	158
2	United Kingdom	367.8019	2493.249441	206
3	France	541.4579	1705.762141	190
4	United States	291.4558	1666.148717	1176
5	Netherlands	331.6250	1640.628288	32
6	Germany	487.8762	1379.571603	105
7	Canada	325.4454	1262.095294	237
8	Ireland	469.4423	990.1880008	52
9	Belgium	480.0000	592.2391219	22
10	Switzerland	301.2619	301.8763585	42

From Table 5.9, we can see that users in France have the largest number of hashtags (on average 541.46 unique hashtags used per user) while the US has the lowest number of unique hashtags per user (291.46). It shows that French users have potentially tweeted a wider range of topics than the users in the US. Germany and Belgium also seem to have a wider range of topics that have been discussed on Twitter (487.88 and 480.00, respectively). Although it is understandable for Twitter users to focus on domestic topics, as many US users do (as well as many users from other countries), in the current dataset, this tendency is more apparent in the US.

Ethnocentrism (e.g., Americentrism or Eurocentrism) topics have been discussed since 1906, and it is common to view things from the perspective of one's own region (Peet, 2005; Hammond and Axelrod, 2006), which is also one of the related factors that this study focuses on to analyse the DH community – environment (i.e., scholars' background). With the development of globalisation, international topics are brought closer to us. Some studies claim globalisation's destruction of the diversity of local cultures, implying that it would create one similar 'American and Western' culture around the world (Holton, 2000; Elteren, 2003), while others believe it can lead to

prosperity among different cultures (Tomlinson, 2009; Machida, 2012). Why do the DH users in the US tweet about a relatively limited number of topics, and many of these topics are domestic-focused? It is an interesting question that requires further analysis.

Nevertheless, the correlation between the number of unique hashtags and the centrality measures can only partly (and weakly) explain the positions in the network that each country has. Other factors also need to be taken into consideration. For example, French users have used more hashtags than users in the UK and Australia, but the latter countries have higher centrality and are located closer to the centre of the network which might be because of the use of English-language. Further investigation is needed to study this question in more depth.

Additionally, the numbers of tweets, retweets, hashtags and betweenness centrality in other countries also raise interesting questions. As mentioned, some countries potentially lose their impact from the number of tweets to the number of retweets, e.g., Ireland and Canada, while some gain more retweets and centrality despite having fewer tweets, e.g., the Netherlands. Due to the small sample sizes, however, this study chooses not to engage further. It should be noted that DH scholars in many countries do not use Twitter as their main communication channel, e.g., China, Japan, Russia (Fiormonte, 2017), and this is a known limitation that will be discussed in chapter 6 (Conclusion).

5.3.6.2 Historical periods

From the discussion in section 5.3.6.1, we know that users in the same country and/or who use the same language tend to retweet and share similar tweets; this is the relationship that has shaped the co-retweet network. However, little is known about how the co-retweet network has been formed overtime. By splitting the time period (2006 – 2017) and visualising networks based on individual year data, longitudinal analysis can assist us in understanding the history of the DH twitter network and its formation.

Based on the longitudinal networks, we can see that organising and holding international DH events (e.g., the ADHO conference) seems to play an important role in new people joining the DH Twitter community. In particular, organising DH conferences in non-Anglophone countries not only brings more local DH scholars to

the community but also encourages non-Anglophone DH scholars from other countries to join in. Below are the Twitter co-retweet networks in each year that are colour-coded with affiliated country information.

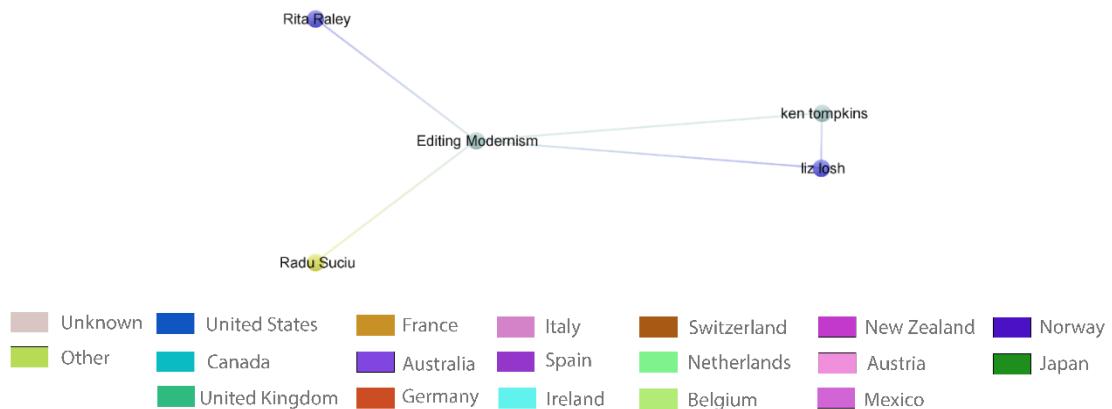


Figure 5.34: User co-retweet network in DH 2009 colour-coded with different affiliated countries, data extracted from Twitter.

Although DH users started to use Twitter as early as 2006, co-retweet networks were only formed from 2009 onwards and with significantly limited representation at first (as shown in Figure 5.34). This is because Twitter only began to offer the retweet function formally in 2009, which then allowed users to share the identical tweet to their followers instead of adding retweet syntax such as 'RT @user' in front of other users' original content. Ross studied three DH conferences in 2009 based on their Twitter hashtags and found that at that time, scholarly activities on Twitter comprised of multiple 'intermittent, discontinuous, loosely joined' conversations. Although there was different retweeting syntax, DH users in 2009 mostly posted original tweets and '@' other users instead of retweeting others' contents, and they would rather comment on the original tweets than spread them (Ross et al., 2011, p. 220). Given the very frequent '@' activities, DH users apparently started to establish connections and interactions on Twitter, although it was still 'fairly small' (Ross et al., 2011, p. 221).

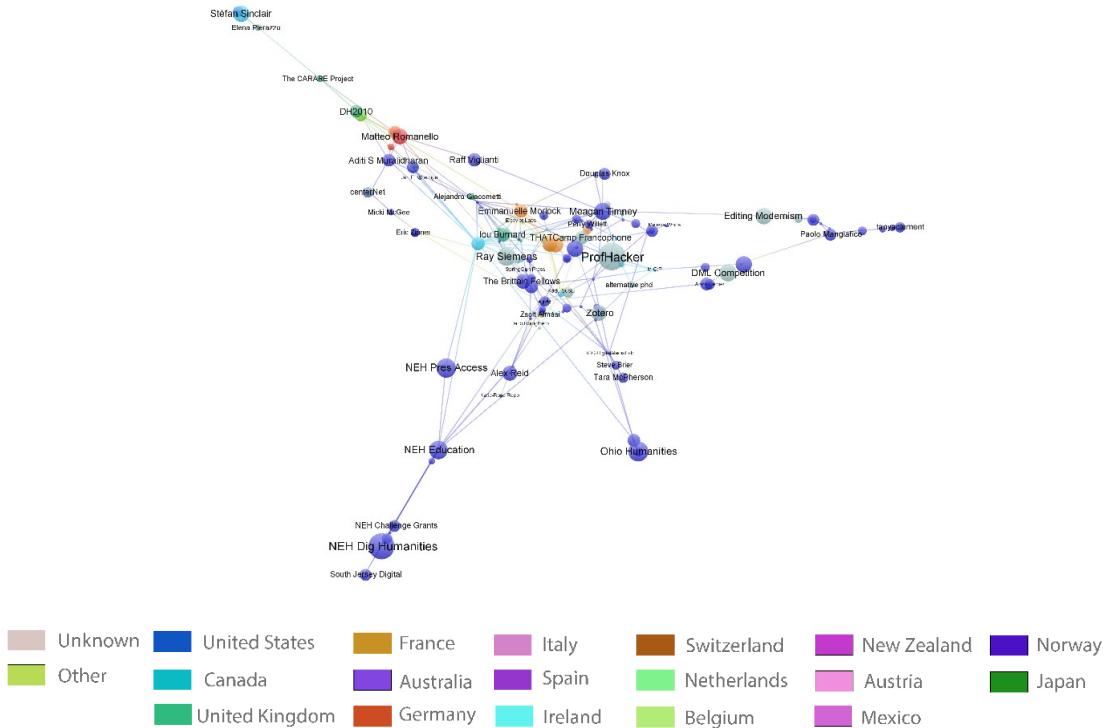


Figure 5.35: User co-retweet network in DH 2010 colour-coded with different affiliated countries, data extracted from Twitter.

The year of 2010 witnessed a significant growth in DH Twitter users from the US and Canada. Organisations, such as 'NEH Dig Humanities', 'NEH Education', 'NEH Challenge Grants', 'Ohio Humanities', created their organisational accounts and joined Twitter in 2010, marking new support from different US institutions (NEH, 2020; Ohio Humanities, 2020) as well as encouragement from Canada, e.g., Editing Modernism (EMiC, 2010). Apart from the US and Canada, there is an emerging group at the top left of the network with newly created accounts related to DH2010, such as 'centerNet' and 'DH2010' in 2010. This is the first time that the ADHO annual conference registered a Twitter account for scholarly communication and also the first time a DH event appeared in the co-retweet network. DH2010 was held in London UK⁶⁷, and around this node, there are many users who were based in the UK and Europe, such as, 'Centre for eResearch', 'lou Burnard', 'Raff Viglianti', 'Alejandro Giacometti', and 'Elena Pierazzo'.

⁶⁷ In 2010, the annual international ADHO conference was held at King's College London by the Centre for Computing in the Humanities and the Centre for e-Research in London, the UK. More information is available at its official website: <http://dh2010.cch.kcl.ac.uk/>.

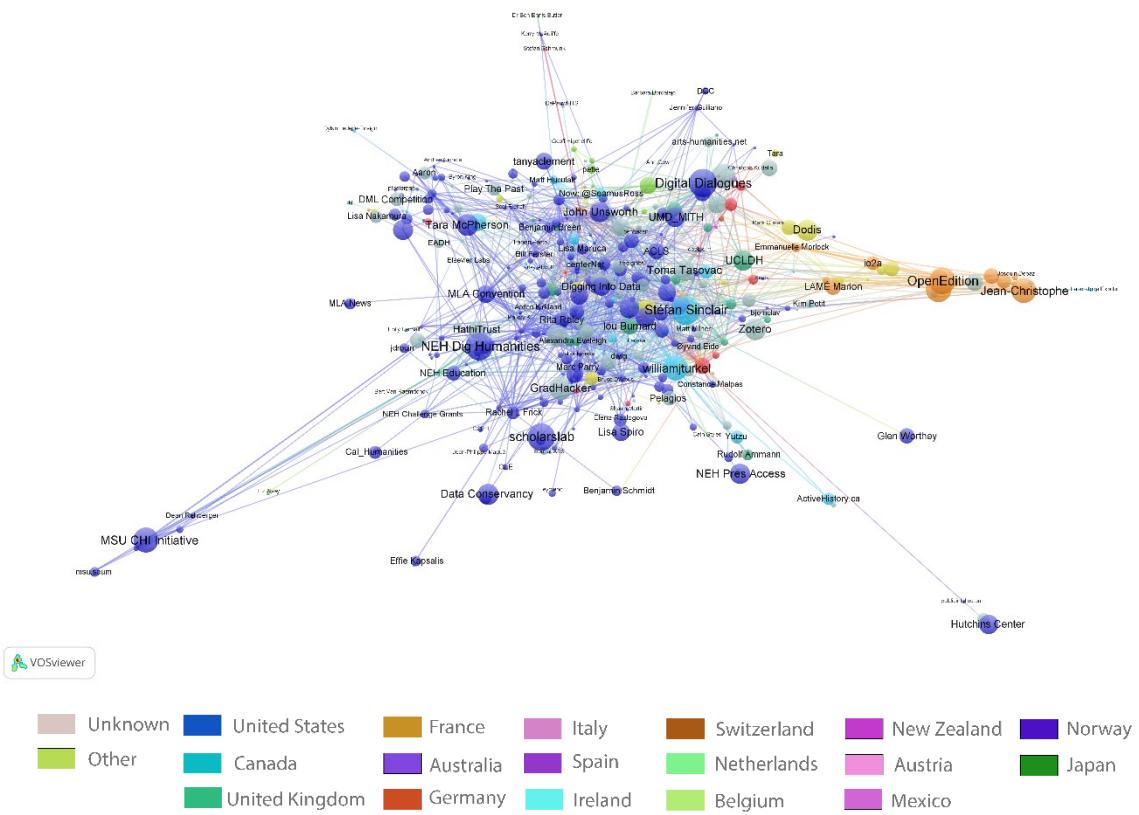


Figure 5.36: User co-retweet network in DH 2011 colour-coded with different affiliated countries, data extracted from Twitter.

In 2011, the number of US users continued to develop and attract retweets. This might be partly due to DH2011 being held at Stanford University in the US⁶⁸. Many of the highly retweeted users are organisational accounts, such as ‘MSU CHI Initiative’, ‘scholarslab’, ‘Digging Into Data’, ‘MLA Convention’, ‘Digital Dialogues’, ‘UMD_MITH’, ‘NEH Press Access’, and ‘Data Conservancy’. There is a group of French users on the right of the network (in orange) that started to form a small but intricately connected cluster. Popular users in this cluster includes ‘OpenEdition’, ‘Aude-Lise Barraud’, ‘Jean-Christophe’, and ‘LAMÉ Marion’.

⁶⁸ In 2011, the annual international ADHO conference was held at the Stanford University in the US, by the University Library. More information is available at its official website: <https://dh2011.stanford.edu/>.

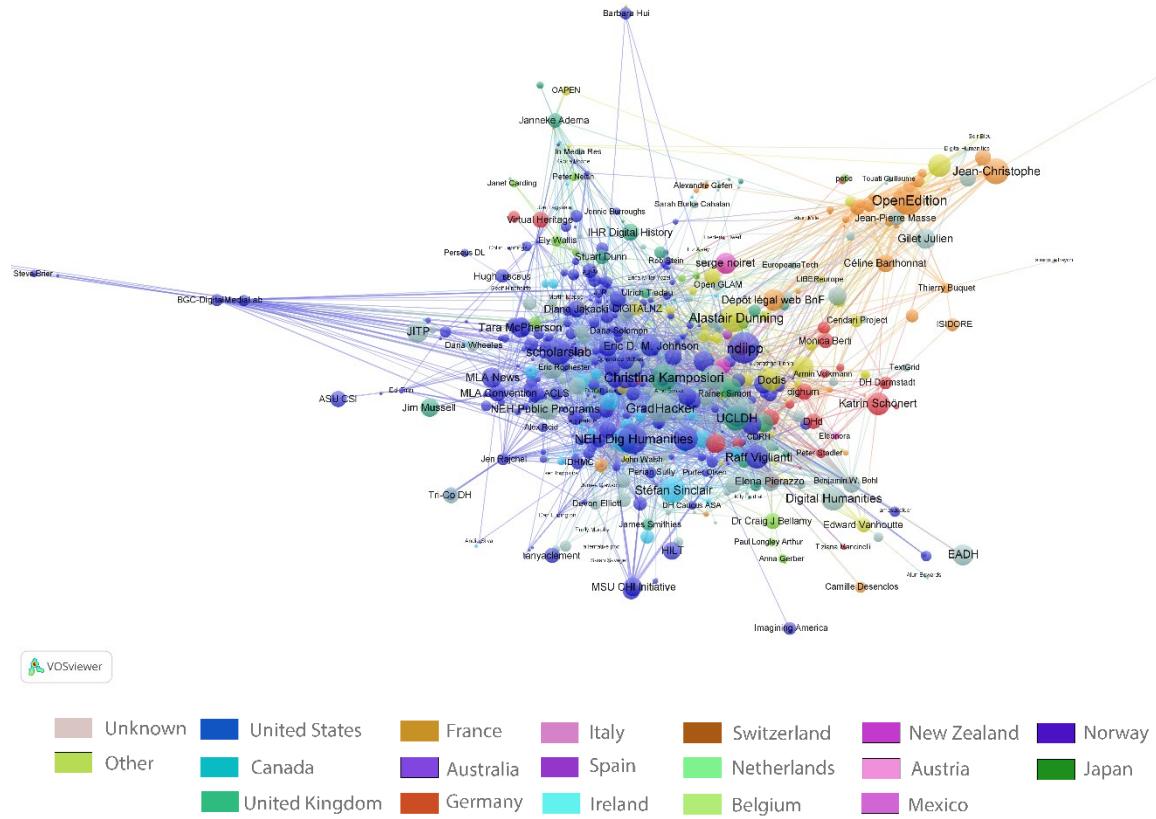


Figure 5.37: User co-retweet network in DH 2012 colour-coded with different affiliated countries, data extracted from Twitter.

As more users joined the co-retweet network in 2012, we can see that apart from the growing number in North American, UK and France, many users appeared in Germany, too. According to the dataset, many of these German users registered their accounts in 2012 which was the year when DH2012⁶⁹ was held in Hamburg, Germany, e.g., 'DHd', 'Katrin Schönert', 'dighum', and 'Peter Stadler'. The number of German scholars (i.e., 42 users) is increased by 162% compared to the number in 2011, while the total number of users only increases by around 58%; it seems that holding the ADHO conferences is particularly important for non-Anglophone scholars to join the Twitter network.

⁶⁹ In 2012, the annual international ADHO conference was held at the University of Hamburg in Hamburg, Germany. More information is available at its official website: <http://www.dh2012.uni-hamburg.de/index.html>.

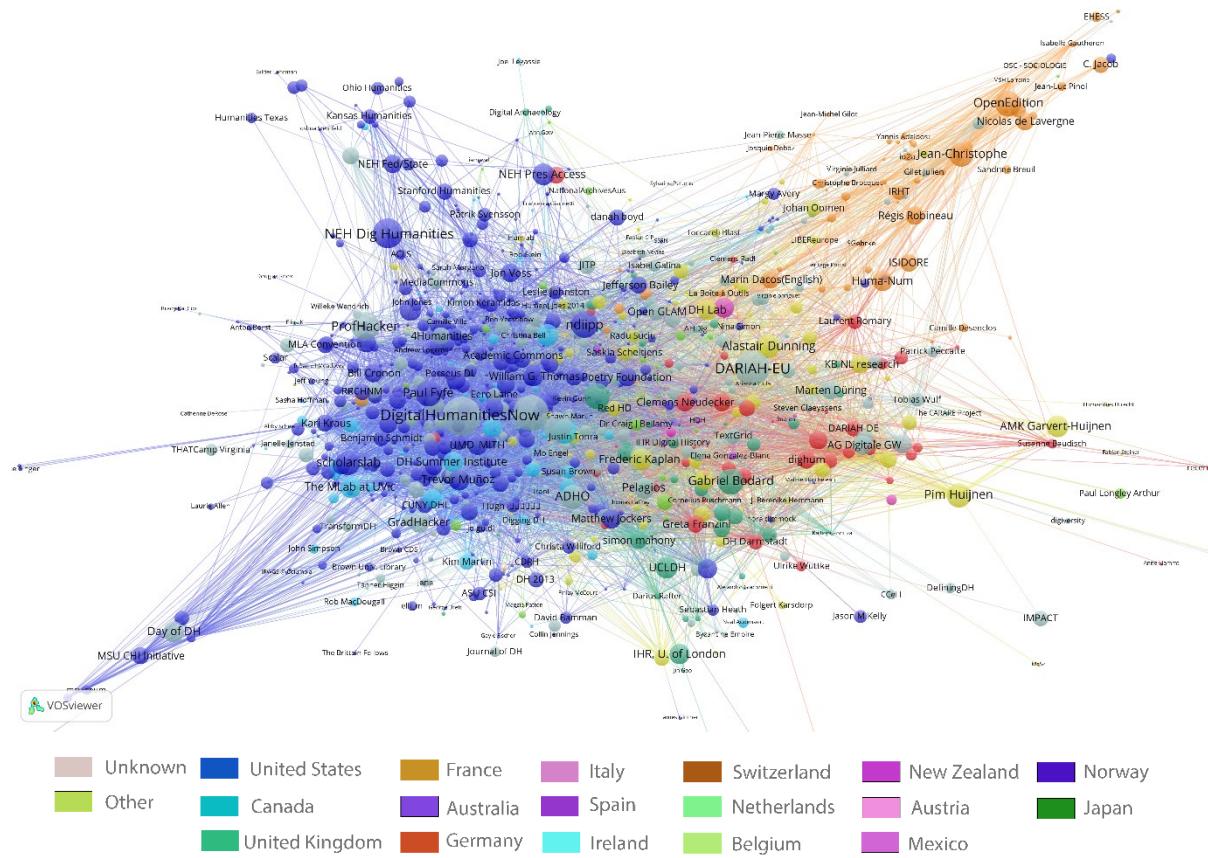


Figure 5.38: User co-retweet network in DH 2013 colour-coded with different affiliated countries, data extracted from Twitter.

In 2013, the number of users continued to expand and increased by around 50% to 1,061 users (from 713 users in 2012). Most of them are from the US, which was the country that held the DH2013⁷⁰ conference. Hashtags such as 'ethics', 'blacklivesmatter' and 'transformdh' were used four times more than in 2012, indicating a growing discussion on diversity related topics.

⁷⁰ In 2013, the annual international ADHO conference was held at the University of Nebraska–Lincoln, the US. More information is available at its official website: <http://dh2013.unl.edu/index.html>.

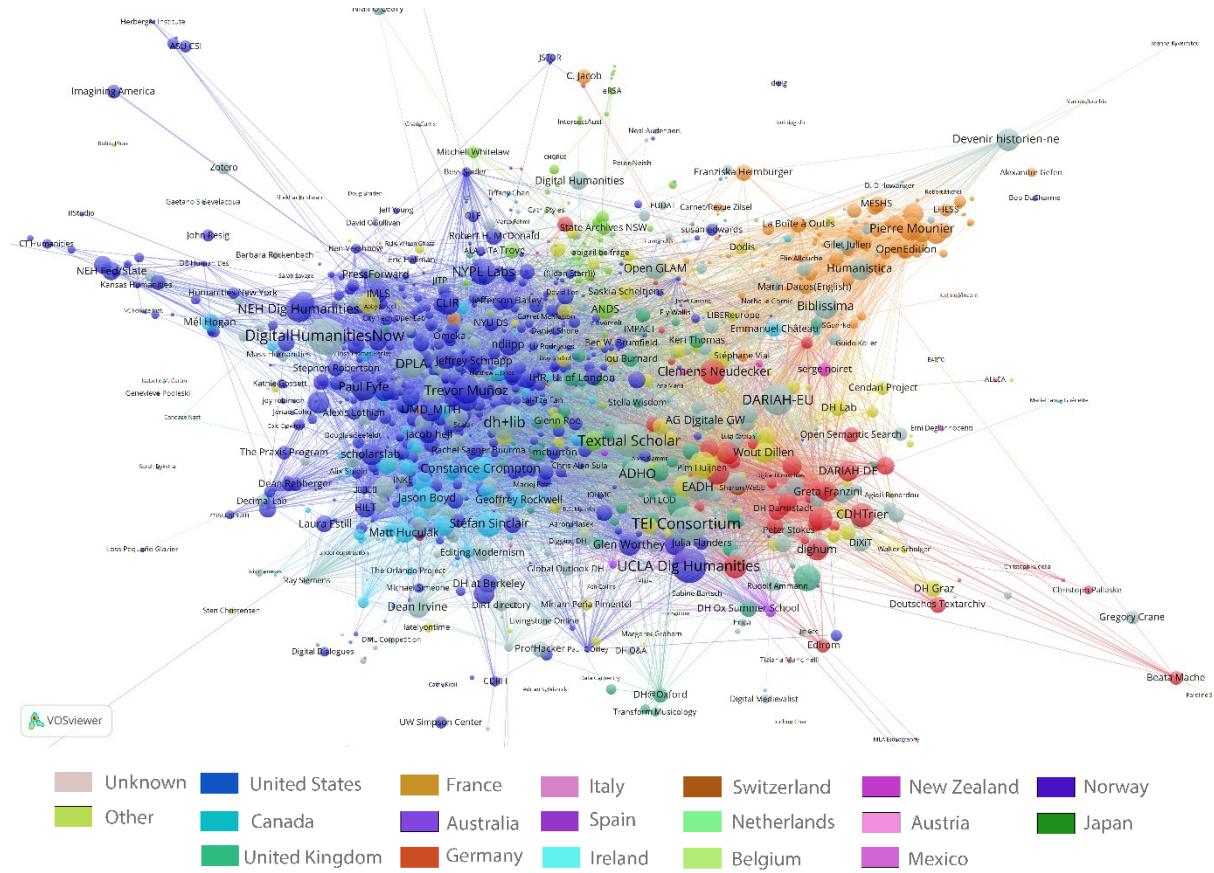


Figure 5.39: User co-retweet network in DH 2014 colour-coded with different affiliated countries, data extracted from Twitter.

The network of 2014 continues to see the DH community expand on Twitter, and non-Anglophone groups such as the French group (top right corner in orange) and German group (bottom right corner in red) are becoming more clustered than before (Figure 5.39). There are 42.21% more users in 2014 compared to 2013, and some countries had more significant increases than others, e.g., New Zealand (166.67%), Japan (100%), Spain (100%), France (67.16%), Belgium (71.43%), and Austria (75%). DH2014⁷¹ was held in Lausanne, Switzerland, a Francophone city, but many non-Francophone countries also witnessed significant growth in user numbers. It raises an interesting question about the correlation between the conference place (i.e., non-Anglophone) and the greater relative increase in non-Anglophone users.

⁷¹ In 2014, the annual international ADHO conference was held in Lausanne, Switzerland by The DBLP Computer Science Bibliography. More information is available at its official website: <https://dblp.org/db/conf/dihu/dh2014>.

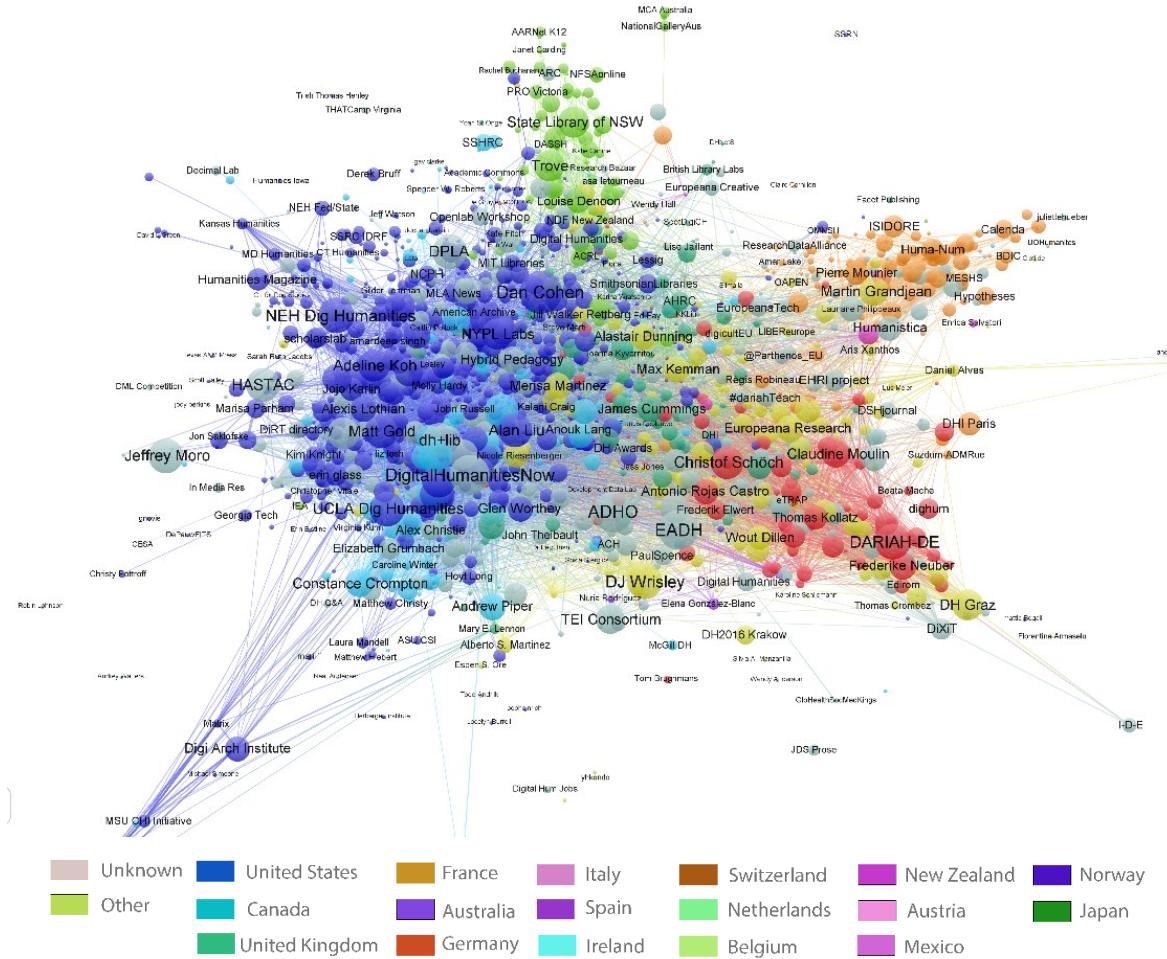


Figure 5.40: User co-retweet network in DH 2015 colour-coded with different affiliated countries, data extracted from Twitter.

Such a correlation pattern can also be found in the 2015 network. In 2015, the ADHO conference was held in Sydney, Australia⁷². As a result, we can see that the cluster of Australian users has emerged (at the top of the network Figure 5.40 in light green) and the number of new users who provided their location as Australia has increased by 74.07% while the general increase in the total number of users in the network is only 26.95% compared to that of 2014. Apart from Australia, there are significant increases in many Asian countries, such as China (85.77%), South Korea (66%), and Japan (50%), etc., and this parallels Weingart's findings based on DH2015 conference

⁷² In 2015, the annual international ADHO conference was held at the University of Western Sydney, Sydney Australia. More information is available at its official website: <https://dh2015.org/>.

abstracts that the proportion of topics on Asian studies nearly doubled compared to the previous year (Weingart, 2014a).

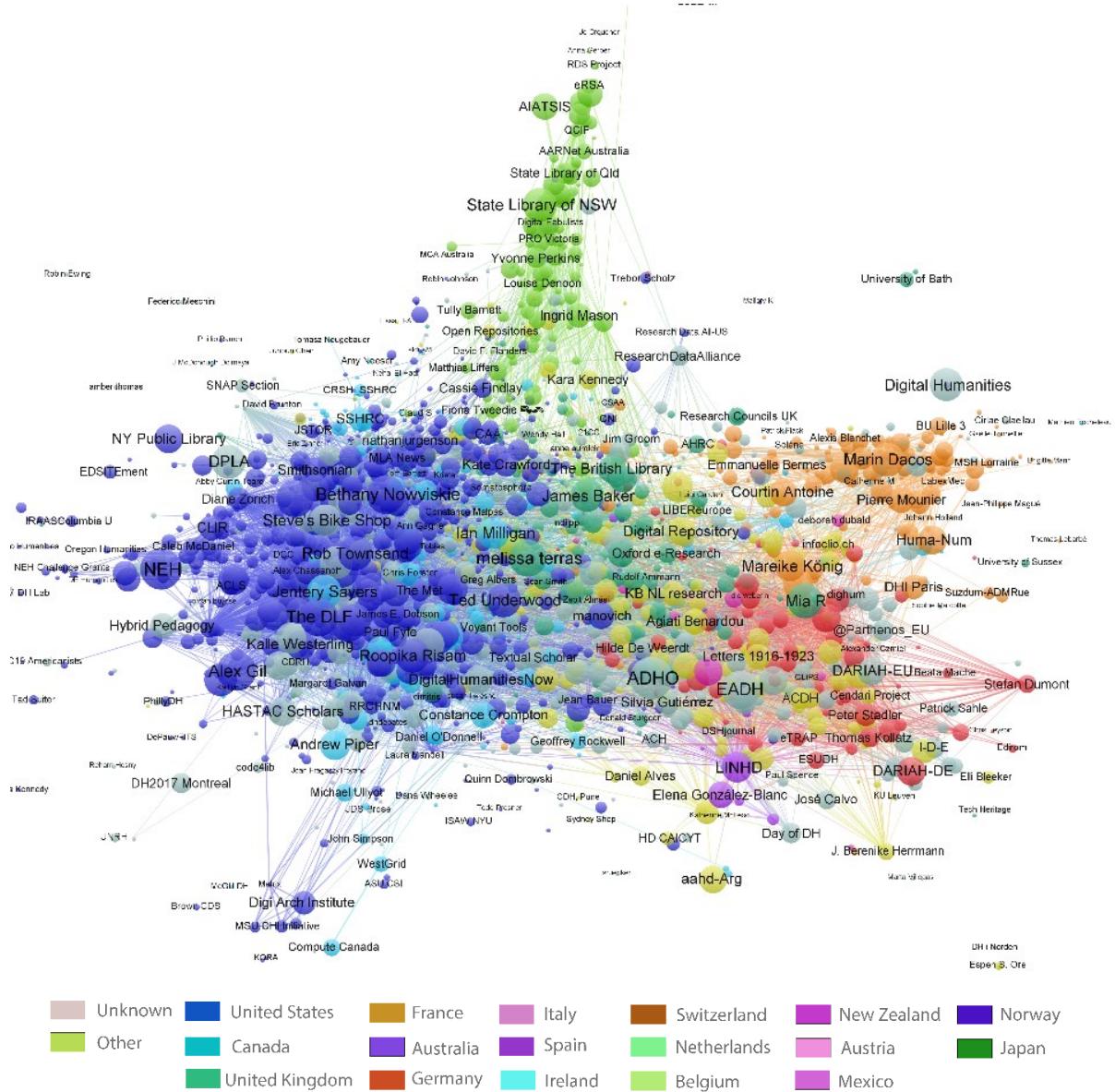


Figure 5.41: User co-retweet network in DH 2016 colour-coded with different affiliated countries, data extracted from Twitter.

In 2016, the Twitter community continues to expand but at a slower pace, and the total number of users in the network only increased by 19.33% compared to previous year. Some countries even have fewer users, such as Switzerland which had a 3.85% decrease in the numbers of users on the network. Some countries, however, show considerable increase in their numbers, such as Austria (57.14%), Ireland (50%), Italy (45.45%), Spain (37.5%) and Mexico (33.33%). Another interesting finding is that the

number of users in Poland does not have a notable increase from 2015 (only 5 users) to 2016 (only 7 users). Given that the DH2016 was held in Poland⁷³, this number is quite surprising when compared to other conference locations. Some suggest that the cost of DH2016 attendance for Polish locals is too expensive (Earhart, 2018; Mahony, 2018), but to investigate the difference in the Polish scholarly use of social media requires further analysis to interpret.

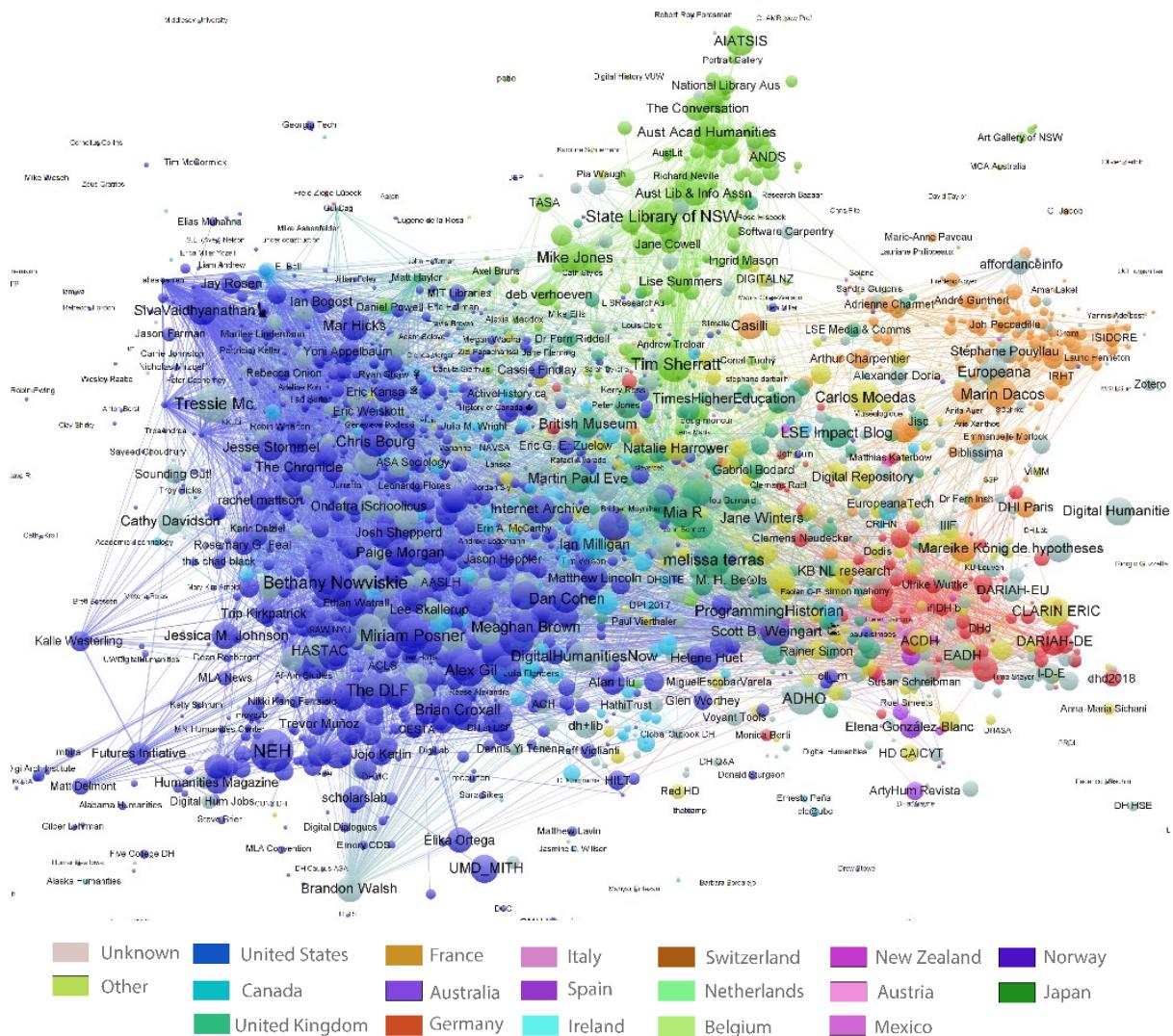
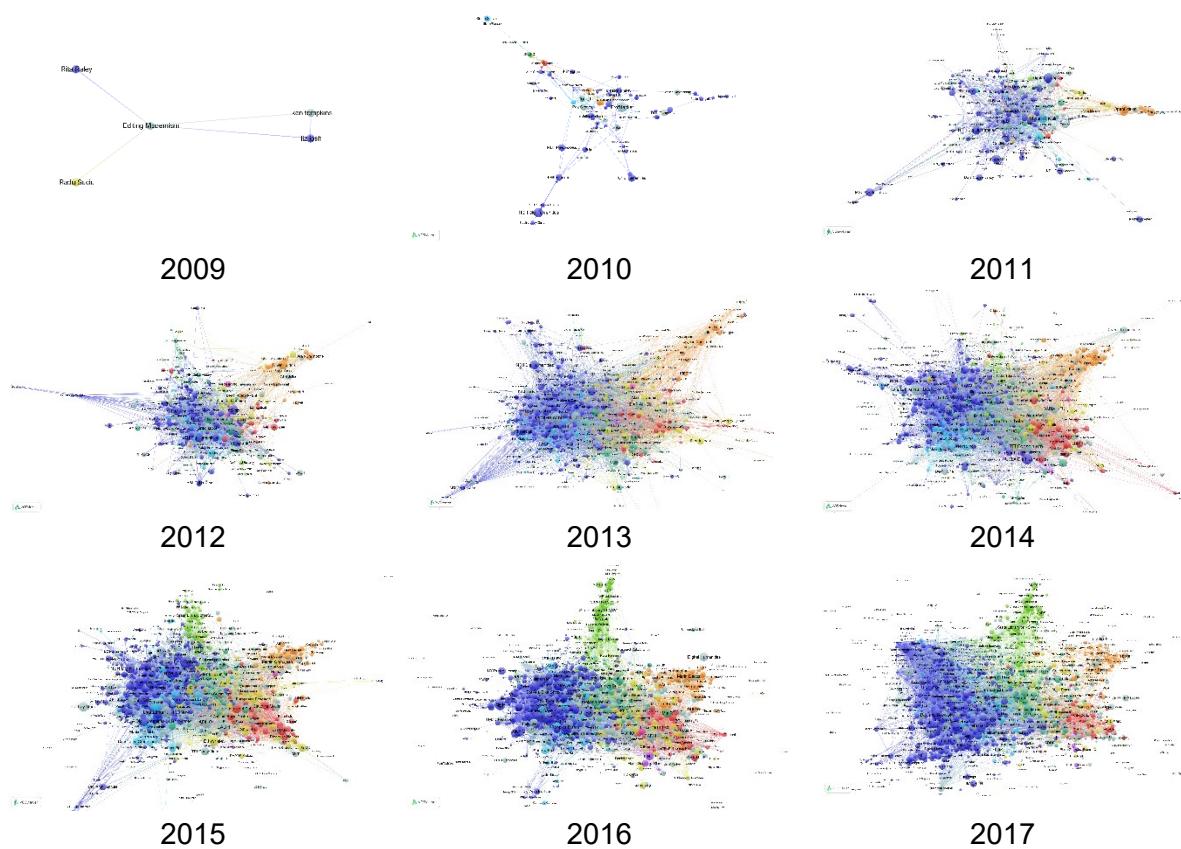


Figure 5.42: User co-retweet network in DH 2017 colour-coded with different affiliated countries, data extracted from Twitter.

⁷³ In 2016, the annual international ADHO conference was held in Kraków, Poland. More information is available at its official website: <https://dh2016.adho.org/>.

The network of 2017 has witnessed the lowest user growth over the whole period (8.9%), and major geolinguistic clusters have started to separate from each other. Compared to 2016, the US, French, Australian and German clusters, in particular, seem to move away and segregate from the centre. The DH2017⁷⁴ host country, Canada, has its users mingled in with different clusters and spread relatively evenly across the network, but it does not have much increase in the number of users, which is similar to Poland in 2016. On one hand, the 2017 network structure and its relative stable number of users for all countries could be a beginning of the network development into a mature status (Tang et al., 2017, p. 985); conversely, this could also be a signal that the DH Twittersphere is closed off to the wider communities and each geolinguistic cluster has appeared to form its own ‘small world’ in DH (Grandjean, 2016, pp. 12–13).



⁷⁴ In 2017, the annual international ADHO conference was held in Montreal, Canada by McGill University and the Université de Montréal. More information is available at its official website: <https://dh2017.adho.org/>

Figure 5.43: DH Twitter co-retweet networks in different years

In conclusion, within the 9-year period that the longitudinal co-retweet network covers, the community has expanded from only 5 users in 2009 to 3,154 users in 2017 and developed into different geolinguistic clusters. Figure 5.43 shows the thumbnails of all-year co-retweet networks and the trend of the community development. As it shows, DH Twitter users started to have co-retweet connections in 2009; they experienced the beginning of multi-regional connection (2010); Anglophonic cluster to emerge (2011); Francophonic cluster to develop (2012); Anglophone to reinforce (2013); Germanophone to emerge (2014); Australian cluster to show (2015); diversity of non-Anglophone countries to increase (2016); expansion to slow (2017). Over time, we can see that organising DH conferences in non-Anglophone countries (e.g., Germany, Switzerland, Australia) not only brings more local DH scholars into the community but also encourages non-Anglophone DH scholars from other countries to join in (only DH2016 in Poland seems to be an exception). Although this dataset does not cover DH2018 which was held in Mexico, it would be interesting to expand the range in the future.

5.3.6.3 Gender

According to the results, the proportions of each gender are relatively even compared to the bibliometric data. In total, there are 1,045 female users (33.16%), 1,097 male users (34.81%), and 1,009 unknown gender users which are usually organisational Twitter accounts (32.02%). Although the number of users in the three categories are balanced, their average number of weights, betweenness centrality, and their tweets vary. Table 5.10 below shows the average value of *betweenness centrality*, node weight (i.e., the number of non-self retweets received), and tweets posted. The complete table can be found in Appendix E.

Table 5.10: the average value of *betweenness centrality*, node weight, and tweets posted in gender on the co-retweet network, data extracted from Twitter, 2006-2017

	average betweenness centrality	average no. nonself retweets received	average no. tweets posted
F	2591.402957	128.271770	1998.051675
M	1405.304172	145.791249	2022.298086
U	918.197033	158.852329	1606.971259

As we can see in Table 5.10, male users posted more tweets than female users. On average, each male user posted 2,022 tweets which is slightly more than that of female users who posted 1,998 tweets on average. It partly explains why male users received more retweets than their female counterpart; as mentioned in section 5.3.6.1 (Country), there is a moderate positive correlation between the number of tweets posted and the number of retweets received (i.e., the more you tweet, the more retweets you will usually get). On average, every male user received 146 retweets and every female received 128 retweets. However, female users occupy more central positions (2,591 *betweenness centrality*) on the network, indicating that tweets posted by female users attract wider interest than their male counterparts (1,405 *betweenness centrality*). This gender difference is similar to that of the co-authorship network (section 4.3.7.2 Gender) where male scholars have higher numbers of publications but lower centrality on the network. It seems that on both networks, male voices are louder but female voices reach further. On the other hand, the users with unknown gender category are distinct from the male and female categories in two ways. Although users in this category tweeted the least, they received the highest average number of retweets. This might well be because the users with unknown gender are usually organisational Twitter accounts, and their tweets are often informational and so attract more retweets. These users also have very peripheral positions on the network with the lowest betweenness centrality values (918 *betweenness centrality*). It is not surprising as organisational accounts usually tweet about news particularly related to that organisation and thus might have a relatively narrow range of topics which do not reach to the wider audiences across different geolinguistic clusters.

Different countries have different gender distributions, and Figure 5.44 shows the gender distribution in the top 10 countries with the most users in the current dataset.

The number of users in genders in the top 10 countries with the most users

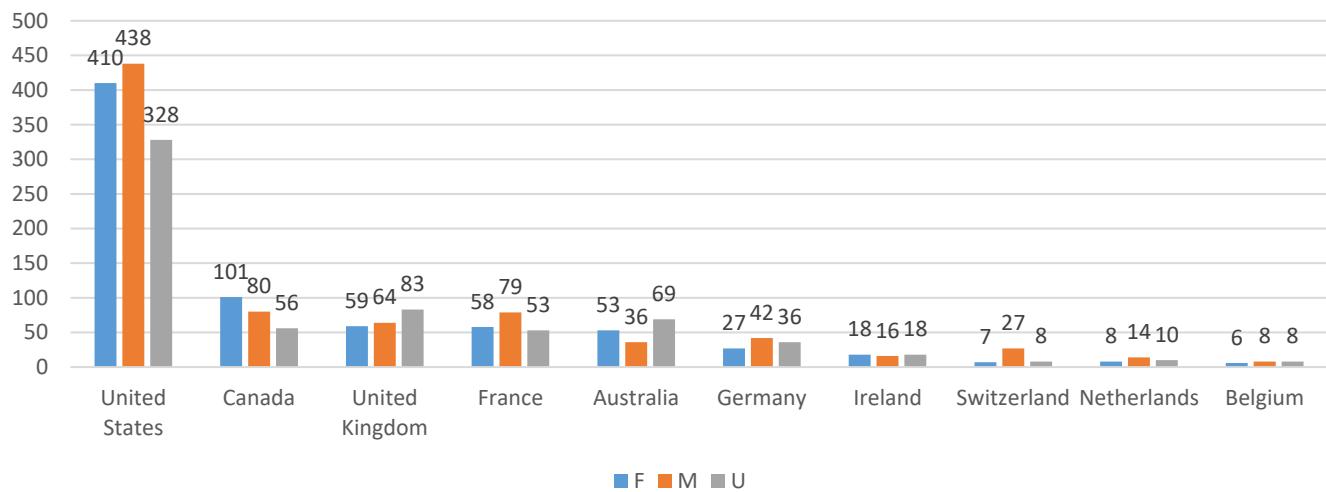


Figure 5.44: The gender distribution in the top 10 countries with the most users in the current dataset, data extracted from Twitter, 2006-2017

Although it is acknowledged that the sample size varies from country to country, we can still learn from the different gender distributions in each country. Some countries have more female users (e.g., Canada, Australia, Ireland), while some have more male users (e.g., the US, France, Germany, and especially Switzerland has 27 male users and only 7 female users). The difference in gender distribution in different countries seems to be connected to the most common topics that female and male users tweet. Table 5.11 below shows the 30 most used hashtags by DH female and male users in the current dataset, and a complete table can be found in Appendix E.

Table 5.11: The 30 most used hashtags by DH female and male users in the current dataset, data extracted from Twitter, 2006-2017

	female favoured hashtag	count	percentage	male favoured hashtag	count	percentage
1	dh	8,294	0.80%	digitalhumanities	9,351	1.01%
2	digitalhumanities	7,933	0.76%	dh	7,682	0.83%
3	dh2017	6,285	0.61%	dh2017	4,863	0.52%
4	dlfforum	5,178	0.50%	opendata	4,764	0.51%
5	twitterstorians	4,757	0.46%	openaccess	4,621	0.50%
6	openaccess	4,203	0.40%	history	4,175	0.45%
7	dhs2017	4,012	0.39%	twitterstorians	3,808	0.41%
8	thatcamp	3,706	0.36%	thatcamp	3,147	0.34%
9	dh2016	3,367	0.32%	highered	2,917	0.31%
10	fb	3,094	0.30%	dh2014	2,882	0.31%
11	history	3,060	0.29%	dh2016	2,881	0.31%
12	dhs2015	2,754	0.27%	envhist	2,839	0.31%
13	dhs2016	2,666	0.26%	dataviz	2,657	0.29%
14	mla16	2,577	0.25%	edtech	2,608	0.28%
15	archives	2,524	0.24%	bigdata	2,592	0.28%
16	mla17	2,438	0.23%	dlfforum	2,199	0.24%
17	highered	2,245	0.22%	iiif	2,176	0.23%
18	libraries	2,075	0.20%	ironème	2,140	0.23%
19	opendata	2,071	0.20%	ironèmes	2,088	0.22%
20	digped	2,041	0.20%	humanities	2,062	0.22%
21	dh2015	1,939	0.19%	digped	2,051	0.22%
22	sharp17	1,890	0.18%	archives	2,034	0.22%
23	dh2014	1,845	0.18%	ai	1,998	0.21%
24	cwcon	1,763	0.17%	fb	1,930	0.21%
25	phdchat	1,738	0.17%	oa	1,908	0.21%
26	archaeology	1,616	0.16%	digiclass	1,876	0.20%
27	dhs2015	1,616	0.16%	dh2015	1,764	0.19%
28	oa	1,602	0.15%	oer	1,758	0.19%
29	musetech	1,472	0.14%	libraries	1,749	0.19%
30	dhs2014	1,452	0.14%	qanda	1,657	0.18%

In total, female users have used 1,038,726 unique hashtags, which is slightly higher than the number of unique hashtags that the male users have used (929,972). As shown in Table 5.11, female users used 'dh' the most while male users used 'digitalhumanities'. It could be that female users tend to pack more information in a tweet (with limited length) but it could also be that female users tend to keep the tweet short and concise. The third favourite hashtag for both genders is 'dh2017' which

indicates the significant influence of ADHO events in the DH Twittersphere for both genders. Starting from the fourth most common hashtag, we can see some contrast in the two genders where females mostly used 'dlfforum' while males used 'opendata'. As mentioned earlier, 'dlfforum' is a forum held by the Digital Library Federation (DLF) in the US (DLF, 2020), and it is the hashtag most used by Bethany Nowviskie (who is the most retweeted user in the current dataset). However, 'dlfforum' is only ranked the 16th most used hashtag by male users, and they seem to be more interested in tweeting about data, e.g., 'opendata', 'dataviz', 'edtech', 'bigdata'. In addition, while common topics posted by females are mostly in English, we can see clear evidence of French-language hashtags among the most used ones of male users, e.g., 'ironème' and 'ironèmes'. This finding corresponds with the significant gender gap in Francophone countries where France has 36.2% more male than female Twitter users and Switzerland (partly Francophone) has 285.71% more male users (Figure 5.44). It also partly reflects the finding in section 5.2.6.3 that non-Anglophone contents are mostly associated with technical topics in the current dataset. In addition, #fb appeared 5,443 times, and it is used by people who use the automatic Twitter update application on Facebook where tweets ending with #fb are automatically imported to Facebook. Female users have included 'fb' many more times (3,094) than male users (1,930), suggesting that female users might be more active on Facebook. This also raises the need for new data sources, as DH scholars might be potentially more active on other social media platforms, as discussed earlier, which could be extended in future studies.

It is shown in the results that certain hashtags do attract one gender particularly more often than another. We have taken six hashtags as examples from the popular hashtags in Table 5.11, and Table 5.12 below is the number of times that these hashtags have been used by each gender.

Table 5.12: Six hashtags as examples used by DH female and male users in the current dataset, data extracted from Twitter, 2006-2017

Hashtag example	Total number of times used by	
	Female	Male
1 #dlfforum	5,178	2,199
2 #history	3,060	4,175
3 #opendata	2,071	4,764
4 #bigdata	1,400	2,592
5 #transformdh	1,281	178
6 #humanities	1,163	2,061

In general, male users have used these hashtags more often than female users, especially 'bigdata' and 'opendata', which attracted 1,000 more uses by males than females. While male users prefer to use more data-related hashtags, female users have used 'transformDH' 1,281 times, which is nearly eight times more than male users (178 times). Female scholars seem to be more interested in diversity topics than male scholars, and this finding also matches Weingart's investigation into DH conference abstracts. As Moravec commented on Weingart's investigation (Weingart, 2015e):

Scott Weingart's investigation of gender at digital humanities conferences suggests, however, that power disparities continue to exist, at least for gender, in these face-to-face events. 'Women are (nearly but not quite) as likely as men to be accepted by peer reviewers at digital humanities conferences' except that 'a lot of the topics women are submitting in aren't getting accepted to digital humanities conferences' including, unsurprisingly, gender studies a field that has almost 70% of its submissions by women, as well as 'culture, teaching digital humanities, creative arts & art history, GLAM, institutions.' (Moravec, 2018, p. 186)

Despite many appeals for more diversity in DH from both genders (e.g., Fiomonte, 2014; Galina, 2014; Mahony, 2018), there has been a trend to devalue topics that women tend to present at DH conferences (Weingart, 2016a). Studies such as gender related topics still seems to be marginalised (Risam, 2015b), and the results have shown that (mostly) only women talk about 'transformDH' related topics.

In general, this section offers a co-retweet perspective to understand the gender distribution and difference in the DH community, and it can help to raise new questions to understand 'who we are' and 'what we can do to improve' in the future.

6 Conclusion

Science, in its broad sense, can be defined as a cognitive network of knowledge as well as being able to be explored as a social network of scholars. Mali et al. explained the difference between the two types of networks as

The cognitive structure of science consists of relationships between scientific ideas, and the social structure of science is mostly manifested as relationships between scientists. (Mali et al., 2012, p. 212)

With the help of the Invisible College research model, this research has quantified, visualised, and interpreted the DH subject and community as four networks. This chapter summarises the four networks visualised in chapter 4 and 5, and addresses the three research questions about DH *subject*, *scholar*, and *environment*.

As can be seen from the results, digital humanities has clearly been expanding and developing in all three aspects that the current study focuses on (i.e., *subject*, *scholar*, and *environment*) during the period studied. However, along with the rise of DH are questions related to its intellectual subject, scholarly collaboration, as well as other important components of its larger academic ecosystem that is arguably transforming academia in significant ways (Gold, 2012, p. ix). The most noticeable question is the distinction between the DH subjects and values we celebrate and their representations in networks. For example, the DH subject network (Figure 4.10) does not seem to include most humanities disciplines where DH has been (or is described to be) rooted. Humanistic topics from only a limited number of disciplines are represented in the DH subject network. Their communication pattern on Twitter (Figure 5.30) seems to form a ‘small world’ while their co-authorship network (Figure 4.15) reflects isolation between groups. Additionally, the visualised communities show a predominantly male and English-speaking geographic distribution that are at odds with the stated diversity values of DH.

Network visualisation is a complicated system, and it can be formed by different elements. Wittgenstein’s ‘family resemblance’ theory was mentioned by Weingart when he discussed the DH keyword network and that a network can be constructed by a series of overlapping forces where no individual force is common to the whole network (Wittgenstein, 1953; Weingart, 2012). Therefore, looking at the cluster

distributions on the visualised networks, it is difficult to explain its formation by a single force. It is also difficult to interpret the reasons for the distinction between how DH is described and how DH is represented in networks.

6.1 Q1 – DH subject

What research topics is the DH subject composed of?

From a bibliometric point of view, DH knowledge is mainly composed of topics related to general historical literacy and information science, computational linguistics, English studies, and early DH projects and pioneers.⁷⁵ These four clusters of topics are loosely connected in the network (Figure 4.10), indicating that scholars are more likely to be interested in just one of those clusters, although it is also noted that some might belong to multiple ones, or it could be argued that they do not belong to any. At its current status, DH as a field clearly contributes more towards topics related to the general fields of historical literacy and information science. Longitudinally, the field was heavily involved in the development of computational linguistics, especially during the three early periods (1966-1970, 1971-1985, and 1986-1990). It is shown that, currently, DH is no doubt an interdisciplinary field, and topics are from a variety of disciplines. Although the interests of these disciplines are not evenly distributed and DH has not yet torn down the walls between academic departments, it certainly has made talking across those walls easier.

From a social media point of view, DH knowledge is composed of four closely intertwined groups of topics: DH and events, methods, non-Anglophone DH, history and other Humanities fields (Figure 5.10). The hashtag co-occurrence network is very dense (98.34% of the hashtags were connected to the main network) with no apparent clustering, and the topical distribution can only be viewed after semantic reasoning is added. Longitudinally, ‘DH events’ has been one of the main forces that formed the network as Twitter performed (and still performs) an important ‘backchannel’ of communication especially during various DH events (Puschmann et al., 2011; Ross et al., 2011). From 2014 onwards, event-related hashtags have gradually moved away

⁷⁵ It is aware that these cluster names are not labelled as any one type of category, as there is no standard taxonomy. Instead, they are named as the most understandable format to better assist the interpretation and discussion.

from the central position to the fringes which indicates an increase in informational use and a decrease in conversational use by DH scholars. It reflects that along with the development of the Twitter community, users have started to become more selective and experienced and perform both conversational and informational tweeting activities (Holmberg and Thelwall, 2014; Myers et al., 2014, p. 498). As the results show, the method-related topics on the hashtag network are more diverse than those found in DH conferences or journals, and there is a clear preference towards the humanistic use, analysis and thinking of digital methods. Although almost all Humanities disciplines can be found in the hashtag network, some have appeared much more frequently (e.g., history, education, politics – including news hashtags especially in North America), and the frequencies of such hashtag usage have been changing overtime.

In general, the Twitter subjects are densely and intricately connected whereas clusters on the ACA network are more loosely linked. To some degree, it is not a surprising result. Research topics are generally more specific and take a longer time to develop and change on publications, whereas users on social media probably share a wider range of interests (including personal and non-academic interests) that are more dynamic and easier to change. The ACA 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.

Because of such different structures, the two networks present different lenses through which to view the DH subject development. The DH knowledge structure seems to be in the process of developing into a mature status on the ACA network, i.e., the process of knowledge integration (Porter and Rafols, 2009; Rafols and Meyer, 2010). On the hashtag co-occurrence network, however, the knowledge integration is still at its early stage, and the topics are densely mixed, with only a few sections beginning to show the initial form of clustering (Türker and Sulak, 2017). The explanation for such difference might be the different time periods that the two networks are based on. While ACA is based on 52 years of data (1966 – 2017), the hashtag network is based on 12 years of relatively recent data (2006 – 2017). On one hand, scholarly subjects

often need time to integrate and mature (Rafols and Meyer, 2010), and 12 years is relatively short for the subjects on the hashtag network to develop and cluster. On the other hand, the ACA network is not yet at a mature stage as its largest cluster (cluster A – the general historical literacy and information science) still consists of various intertwined subjects (e.g., history, information science, literacy, new media). This cluster A is the newest cluster on the ACA network and almost all its topics can be found in the hashtag network, while other older clusters on ACA can hardly be found and matched on the hashtag network (e.g., the cluster B of computational linguistics), which are believed to have split from the DH realm and moved to other fields (Jensen, 2014; Mitkov, 2014). In other words, if we only look at the recent periods (2006 – 2017), the subjects on both networks are generally overlapping despite the fact that the two groups of people (the subjects) are not the same. Although the overall time periods of the bibliometric and Twitter data are vastly different, nevertheless, they could still provide a complementing longitudinal history told as stories of bibliometrics and social media use.

Longitudinally, DH history has witnessed the field's development from counting words to studying all types of media and employing a variety of methods. Both on publications and on social media, the field is 'no longer an ivory tower' (Burdick et al., 2012, p. 82). Especially on Twitter, DH scholars discuss not only Humanities topics and digital methods but also news, policies, and current movements. It has experienced the disciplinary transformation over time from engaging in traditions to accommodating growing capacity, diversity, outreach and engagement (Siemens, 2016, pp. xxi–xxii), and this study has provided quantitative and visualised evidence of such disciplinary change.

In addition, in both networks, the technical related sections (i.e., computational linguistics, methods) are on the edge of the networks and are not densely mixed with other sections. A close and strong connection between non-Anglophone communities and the technical aspects of DH is found in both networks, indicating an existing language barrier for non-Anglophone scholars when using English to discuss Humanities topics (instead of technical topics) in DH whether in publications or on social media. This is an important finding of this study that can be brought to our attention, and most DH scholars do not seem to have acknowledged or discussed it

extensively yet. Whether to ‘yack’ or to ‘hack’ should not be an activity that is heavily influenced by the language one speaks, and everyone should be able to conduct and present their research without linguistic barriers. As part of the community, we should not only encourage the geo-lingual diversity both in publications as well as on social media but also review what we present in English. Do we ‘yack’ too much and discuss too many Anglophone country-oriented regional topics in English? In order to integrate the ‘digital’ and the ‘humanities’ in DH, we also need to address new ways of understanding and balancing the geo-lingual topic preferences.

Network visualisations help us see a quantitative aspect of the DH subject, and to further understand its subject history as well as its future, one needs closer investigations on the networks and to analyse previous DH scholarship in more depth. Many studies have suggested a growing trend of more diversified topics, and that the connections between DH and the humanities are growing stronger, e.g., (Weingart and Eichmann-Kalwara, 2017). After seven decades, the stakeholders of DH ‘remain committed to bringing together technology and the human record synergistically’ (Poole, 2017, p. 94). ‘Neither the traditional nor the digital humanities can succeed alone as well as they can together’ (Hayles, 2012, p. 61), and the future is ‘up to individual scholars to respond’ (Terras, 2016, p. 1645).

6.2 Q2 – DH scholar

Who has contributed to the development of DH? As this study tries to avoid identifying individuals (except when giving examples), it focuses more on addressing the research question by revealing the social patterns of the DH communities (i.e., collaboration and Twitter sharing interaction).

From the bibliometric point of view, only a small proportion of DH scholars (37.99%) have co-authored, and only 19.54% are connected to the co-authorship network, indicating a ‘small-world’ collaborative model where everyone knows everyone. The collaborative nature of DH, as accepted by many (Deegan and McCarty, 2012; Flanders, 2016), was not found during the early days of this field, or at least this aspect of its nature was not shown in publications. It has, however, developed throughout the expansion of DH and has become increasingly notable in the past decade (e.g., 69.49% in 2013, 66.07% in 2015, and 70.24% in 2017). However, DH is not unique in the

growth of scholarly collaboration, and it is here as elsewhere partly due to the wider academic context, i.e., the improvement of communication channels, the policy-making preference of funding groups, and journal editorial decisions, etc.

From the social media point of view, we can see that the DH co-retweet network on Twitter is formed mainly by language (based on the user interface language setting) and location (based on the public user profile). In other words, users who are in the same country and/or speak the same language tend to retweet and share similar tweets. To some degree, the infrastructure of Twitter (e.g., recommendations and trends for you) is playing an important role in the process of disciplinary community formation and cohesion (Hashemi, 2017). How we choose and use social media also influences the building of such infrastructure and in turn enhances the monolingual clustering on Twitter (Sehl, 2020; Twitter, 2020a, 2020b). In addition, it is interesting to notice that most of the highly retweeted tweets in the current dataset have included links, demonstrating a strong informational use that potentially plays an important part in forming the DH social network. Longitudinally, not all countries and language groups joined the network at the same time. The results show that organising and holding international DH events (e.g., ADHO conference) seems to play an important role in people from geolinguistic groups joining the DH Twitter community and taking part in the popular discussions.

In general, the formation of both DH social networks (co-authorship and co-retweet) seems to be heavily influenced by non-academic and non-intellectual factors, such as language, working country and institution, as well as informal relationships. It parallels what Ponomariov and Boardman once pointed out in their scientometric study:

The results suggest there are numerous dimensions of co-authorship, the most influential of which is informal and relational and with little (directly) to do with intellectual and/or other resource contributions. (Ponomariov and Boardman, 2016, p. 1939)

Such formation also corresponds to the very idea of an original ‘invisible college’ that was first coined by de Solla Price (De Solla Price, 1963; De Solla Price and Beaver, 1966); this has been revised and applied as the methodological framework in this study. When de Solla Price first proposed the original idea of the ‘invisible college’, he defined it as a small group of approximately 100 people and emphasised the informal

relationships between scholars as being crucial to knowledge production (De Solla Price and Beaver, 1966; Zuccala, 2006). Many have agreed that the process of knowledge production, especially in the Humanities, is community-driven through the use of footnotes and other humanistic ways of information collection (Brockman et al., 2001; Earhart et al., 2020). It could potentially encourage personally formed connections, restricting the size of the community, and even disproportionately impact on the diversity of scholars and the representation of the communities (Risam, 2015a; Mahony, 2018; Earhart et al., 2020)

Moreover, both networks seem to be at the early stage of development. As mentioned, the DH co-authorship network presents a ‘small world’ with only a small but strong set of them collaborating actively (19.54%), and the proportion is comparatively lower than for other disciplines. For example, the co-authorship network in Medicine contained around 92% of its total authors (Newman, 2001). In Management and Organization, the number was 45% (Acedo et al., 2006), and in Sociology, it was around 34.5% (Moody, 2004). Nevertheless, with the rapidly growing trend of DH co-authored articles in recent years, the network is expected to expand and develop. Similarly, the co-retweet network agrees with the Nielsen ‘90:9:1’ rule, i.e., 90% of users are ‘lurkers’, 9% of users contribute from time to time, only 1% participate actively and account for the majority of contributions (Ross et al., 2011, p. 221). On the co-retweet network, only 77 users (2.44%) were retweeted more than 1,000 times (non-self), while 339 users (10.75%) did not receive any non-self retweets, and 2,586 users (82.00%) only received less than or equal to 500 non-self retweets (Figure 5.28 and Figure 5.29). This power-law-like result indicates that retweets tend to come from a relatively small group of original tweets (Ediger et al., 2010), reflecting a significant one-to-many dissemination pattern, although some many-to-many patterns were also identified.

Both network structures (‘small world’ and ‘power-law-like’) can potentially lead to the Matthew effect of accumulated advantage, i.e., ‘the rich get richer and the poor get poorer’ (Gladwell, 2008). The longer scholars progress up the academic ladder, the more accumulative influences they tend to have (Rørstad and Aksnes, 2015), such as the ‘cumulative advantages of academic capital’ (Merton, 1968). This is a vivid example of the social dynamics that such factors could cause, and it requires our

further attention. As scholars who are doing DH related studies, we could also improve the representation by connecting and working with people outside of our ‘small world’.

Nevertheless, it should be noted that independent work should also be acknowledged. Collective working delivers different expertise and ideas, but it also introduces coordination responsibilities and social networking tasks that might make the process of knowledge production less efficient (Guimera et al., 2005). Scholars such as Wuchty et al., quoted Fitzgerald’s claim that ‘no grand idea was ever born in a conference’ (Fitzgerald and Wilson, 2009) to raise the importance of individual working as:

A shift to teamwork may be a costly phenomenon or one that promotes low-impact science, whereas the highest-impact ideas remain the domain of great minds working alone. An acclaimed tradition in the history and sociology of science emphasizes the role of the individual genius in scientific discovery. This tradition focuses on guiding contributions of solitary authors, such as Newton and Einstein, and can be seen broadly in the tendency to equate great ideas with particular names, such as the Heisenberg uncertainty principle, Euclidean geometry, Nash equilibrium, and Kantian ethics. The role of individual contributions is also celebrated through science’s award-granting institutions, like the Nobel Prize Foundation. (Wuchty et al., 2007, p. 1036)

Collaboration as well as social media interaction can also introduce challenges by many other factors:

Different patterns of work activities, expectations, personal beliefs, specialized language, and individual goals make it difficult for participants to collaborate, explore, and share one another’s specialized knowledge. These differences can cause team members to contest or challenge one another’s contributions, although these differences may also enrich collaboration. (Hara et al., 2003, p. 953)

Communication does not always equal connection, and sometimes, it brings the opposite, especially on social media. It is often reported that outspoken scholars such as Professor Mary Beard are trolled and harassed on Twitter (BBC, 2013; Silverman, 2013; Vince, 2018), and some even used the term ‘Twitter war’ to describe similar ‘nasty’ debates and arguments, in which DH as a field has also been involved (Hagmann, 2018). Meanwhile, some see it from a more amicable perspective:

[...] DH will have to let go of our ideas of niceness and methodological agreement, and accept the likelihood that different schools and methods of doing DH will emerge. This may entail public battles, schisms, and regroupings, but it does not necessarily threaten the integrity of the discipline; it may even be a sign of strength and confidence. (Warwick, 2015, p. 549)

Overall, scholarly social activity is a complex process that is influenced by various factors, and it cannot be simply explained by the patterns of communication and collaboration. There is no right or wrong about working as a team or alone, and DH as a scholar community is 'never conceived of as a homogenous entity anyway' (Terras, 2016, p. 1645).

6.3 Q3 – DH environment

How diverse are the backgrounds of DH scholars (i.e. gender, working country)? How do gender and affiliated country diversities influence the DH *Subject* and *Scholar* structures?

From the bibliometric perspective, we know that the environmental factors (e.g., gender, language, and affiliated country) have played important parts in the formation of DH scholars' co-authorship network.

Although male scholars have dominated the field (66.62%), female scholars (28.86%) have experienced a rapid growth in numbers during the last 20 years, and they have been acting as critical bridges in forming collaborative links. The average betweenness centrality of female scholars is 41.16% higher than that of the male scholars (10,646.99 – female, 7,542.46 – male), and women in DH have relatively more collaborators on average than men (2.62 – female, 2.54 male). The results have shown that female scholars have encouraged more communications, built more collaborations, and contributed more to the formation of the DH co-authorship network than males. They are not only the main forces to maintain the DH scholarly connections but also the icebreakers to bridge and connect isolated groups.

Authors affiliated in Anglophone countries are the majority in the author pool. While the US scholars accounted for a large proportion of the publications, scholars in the UK and Canada were, on average, more likely to contribute to the formation of co-authorship links than any other countries both as bridges as well as central nodes.

Other countries (e.g., Spain, Finland, Germany, Netherlands) have also played important roles in the formation of a global DH network. Despite the growing appeals for more diversity and openness in DH (Galina, 2014; Fiornonte, 2017; Mahony, 2018; Earhart, 2018), this study has found that the level of international collaboration in DH is more extensive than many other disciplines, indicating an ever-growing international collaborative community. On the other hand, it is interesting to notice the considerable gap between the co-author rate and the international co-author rate in some countries. For example, scholars who work in countries such as the US, Japan, France, Switzerland, and China collaborate heavily but often only within their domestic range.

From the social media point of view, although male and female scholars account for similar proportions (female – 33.16%, male – 34.81%), most central positions in the clusters are taken by female scholars (as well as organisational accounts that have unknown gender), who have been acting as critical bridges in forming the co-retweet network. On average, female users tend to discuss a wider range of topics than male users, and these topics are relatively more understandable to many. Conversely, although male scholars tend to receive a higher number of retweets, they are generally positioned further from the centre. In addition, there is a possible connection between gender and their countries' positions on the co-retweet network that has the potential to be analysed further. For example, some countries have a lower female than male rate, and their values of betweenness centrality are relatively lower (e.g., the US and France); while some countries have more female users than male, and their nodes are positioned much closer to the centre (e.g., Australia).

It is clear in the co-retweet network that users using the same language tend to cluster, and Anglophone countries make up the majority (82.94%) of this network, although the number of users based in other countries has also been growing, e.g., the Francophone (7.93%), the Germanophone (3.77%), the Hispanophone (1.67%). Different Anglophone countries form social connections differently, and users located in Australia and the UK (although they only account for 6.90% and 5.30% of the total users, which is far less than the US, 39.44%) have the highest (and second highest) average weight (i.e., number of non-self retweets) and betweenness centrality (i.e., the ability to bridge different nodes). This reflects that a large number of users from different countries and language groups have retweeted the tweets posted by

Australian and UK scholars, and this is partly because users in these two countries have tweeted more often about various global DH topics rather than only local domestic ones.

In general, we can see that in both networks, female scholars in DH are playing relatively more important and central bridging roles to form both sets of communities, although their proportions are different (female/male on co-authorship network is around 1:2, while it is 1:1 on the co-retweet network).

Language and country are also shown to be critical determinants of the DH community formation, and the Anglophone communities dominate both networks (Flanders, 2016; Pitman and Taylor, 2017; Tello, 2017). It is not unexpected as both the publication data and Twitter data have been collected mostly from English-language sources. On one hand, language, as an archival textual material, is important in DH:

DH invites contribution from all Humanities disciplines, including those where language plays a secondary role, such as anthropology, archaeology, fine (visual) arts, film studies, and musicology. These are not the most likely disciplines for computational linguists to get involved in, but linguistics and literature (studies) are also Humanities discipline, where language is central, as are history and philosophy, where language is not of central interest, but where archival material in textual form often plays a central role. (Nerbonne and Tonelli, 2016, p. 7)

On the other hand, language, as a means of communication, is also a significant barrier to the community formation, and many have raised concerns about the diversity of the DH scholars. Some pointed out that DH had become increasingly 'exclusive' and 'cliquish' (Bianco, 2012; Edwards, 2012; Pannapacker, 2013), and others tried to trace the reasons for such misrepresentation and suggest other data sources for studying the DH community representation (Mahony, 2018; Risam, 2018). ADHO official journals and Twitter connections indeed provide relatively authoritative data, but they are known to be highly Anglophone-oriented. Fioronte discussed the 'anxiety and fear of being "cut out" from the "international" game' and explained the formation of the ADHO Steering Committee:

The real issue was ADHO, an organism that defines itself as internationally representative of the Digital Humanities, but that still lacks a bottom-up

democratic structure. The members of the Steering Committee are not elected by the members, but by the boards of each Constituent Organization. The reason is that ADHO was created by a club of 'constituent organizations' (USA, UK, Australia, Canada and Japan), which, in fact, gets to decide the who, how and why of membership. [...] While with these new changes ADHO is struggling to sell an 'international' image of the community, most intellectual tools remain in the hands of the Anglophones: the annual conference, the Humanist mailing list, the monolingual LLC/DSH journal, the more or less sponsored monographs (such as the Companions). Not to mention software, languages and so-called 'standards' like the Text Encoding Initiative. (Fioronte, 2017, p. 6)

Moreover, public information on social media and publications can lead to denser monolingual social connections because of them being public. Similar to the Mathew effect, it will make the community less diverse, as Grandjean explained:

The fact that the list is public is likely to skew the results of this analysis: conscious of having been added, some users could use it to discover and follow new users, which would have the effect of increasing the network's density. [...] In the longer term, a public list is problematic because it is likely, gradually acquiring the status of 'reference', to encourage compulsive subscription behaviours, such as users hoping to be 'followed back' by colleagues. [...] But on the other hand, keeping this list public is mostly a way of giving the community a chance to discover unknown profiles and is a contribution to the friendly spirit of this social media. (Grandjean, 2016, pp. 12–13)

Because of different cultural, political, and social structures, scholars in different regions and/or that use different languages tend to have communications and present scholarship via different platforms. It is clear that English-language journals and Twitter are not the only option to study the DH formal and informal communities.

Some believed that 'the share of each country's academic publication is very close to its share of the world's wealth (measured by GNP)' (De Solla Price, 1986, p. 142). However, scholars affiliated in the BRICS association (Brazil, Russia, India, China, and South Africa) that represents 25% of the world's GDP and 43% of the world's population are hardly seen in ADHO journals, events, and Twitter discussions (Fioronte, 2017, p. 3).

As the candidate noted, there was not one scholar affiliated in Mainland China (the world's second largest economy and with the most population) who attended the DH2018 held in Mexico (where 340 papers were presented), and through personal communications, many Chinese scholars expressed their frustration that their abstract submissions were rejected (many receiving low review scores for poor English). Many Chinese DH studies are published in Chinese-language publications and proceedings, and although the Chinese DH community has been growing rapidly, they hardly use Twitter to communicate, but rather, WeChat and Weibo (Mahony, 2018; Mannion, 2018). Studies have found that the Chinese community is unique in its gender distribution, and female scholars always dominate the field at its biggest annual DH convening (i.e., the Peking University Digital Humanities Forum) with around 61.54% female scholars and 38.46% male scholars (Zhu and Nie, 2016, 2017). By bringing together more non-Anglophone DH communities, like the Chinese one, we could not only improve the geolinguistic and gender diversity but also potentially study their scholarly social structures and learn from each other. 'Restricting linguistic and cultural perspectives restricts our field, whereas inclusion benefits us all.'(Mahony and Gao, 2018)

7 Reflection and future study

This chapter discusses the noted limitations and possible future studies. Although it is impossible to compile a completely representative dataset (either based on publications or social media), limitations can be reduced by adding a wide range of data sources and employing improved methods. Given that this study produces a large number of network visualisations, it is made available openly so that many findings and questions can be unpacked and investigated further in the future.

7.1.1 Importance

This PhD study has the potential to benefit relevant DH studies as well as those of cognate and other disciplines for its multi-dimensional network visualisations, largest open database, and innovative research framework.

It provides visual images of the discipline (both as a whole and at individual level) that contribute to the ongoing debates about the DH knowledge structure, scholarly communication and connection, the formation of the field, and offers a new perspective

to revisit DH publication and social communication. Historians can use the longitudinal networks to dig into other 'hidden' perspectives of the DH history, and newcomers can learn the community structures and communication patterns through direct visual representations.

This research has compiled the largest DH bibliometric dataset that the candidate is aware of, from the three most important ADHO journals that would not be able to collect by using general citation databases. It has complemented the dataset by adding previously missing metadata and building a more comprehensive database that will be released open access in the UCL institutional repository after the completion of this PhD. This database offers not only the bibliometric metadata, but also more detailed information such as gender (by name) and affiliated country (when published) that could be used to further study the contexts of the community structure. It will be a valuable data source and avoid the need to duplicate data collection for other DH studies. It is stored as a CSV file that can be easily processed by software and edited and extended with new data to update the results.

The revised and developed methodological framework used in this research can also be replicated by other projects. This study not only adopts the idea of an Invisible College to study the DH community structure, but also extends its model procedure with new and robust approaches. Because the Invisible College model was originally proposed in 2009, when scholarly communications were mainly via traditional mechanisms (e.g., emails and meetings), some of its original methods have now become limited. For example, stages such as *the social actors* and *the IUE* that investigate the informal communications and background contexts of scholars can nowadays be studied based on social network analysis instead of the conference co-attendance that was planned originally. This thesis enables this research model to go beyond the traditional disciplinary research system by adding dynamic types of data, updated methods, and cohesive procedures. In this way, the new Invisible College research model becomes more compatible with current scholarly communications and more flexible for other disciplines to study their structures and histories.

Furthermore, the output of this study includes an interactive webpage of DH scholar networks that enables the audience to search and filter the names of scholars within

both the citation⁷⁶ and Twitter⁷⁷ networks. This output will serve as a research tool for people who are interested in the DH community to study individual cases from both bibliometric and social perspectives. The visualised networks and results have the potential to be extended to other projects, such as DH genealogy study and DH historical archive study.

To the candidate's knowledge, this study is the first DH network study that employs a well-grounded research model and investigates the DH communities through both formal and informal perspectives. Scholars could take this study as a reference when planning future publications, collaborations as well as scholarly connections. Universities, centres, journals, and organisations might well take the gender and country differences into account when encouraging authors to communicate and publish. Future network studies can also replicate this study in other fields and topics, and this might make for good comparisons between different disciplines.

7.1.2 Limitations

As noted throughout the thesis, there are some limitations that need more attention. This section will focus specifically on the limitations of the dataset representation (7.1.2.1) and the network visualisation (7.1.2.2).

7.1.2.1 Dataset

Although this study aims for more inclusiveness, it is impossible to cover all scholars in the DH field. The two datasets in this study are the result of an ADHO-centric investigation for identifying the respective publications and the Twitter users. ADHO official journals and related Twitter connections offer a useful starting point, but not all DH organisations interact with ADHO, and not all DH communities are on Twitter connecting with ADHO and its member organisations (Galina, 2014, p. 311). Both bibliometric and Twitter datasets are mainly focused on English publications and conversations in English-speaking regions although a small number of global DH communities can also be detected. In addition, other related areas that are at the

⁷⁶ Interactive ACA network: http://jin-gao.com/map/view.html?citations_all

⁷⁷ Interactive co-retweet network: http://jin-gao.com/map/view.html?twitter_all

fringes of the ADHO remit have also digitally matured, and they are useful data sources for future studies, too.

Research funding is one of the most important factors that encourage publications and collaborations (Shewan et al., 2005), and many DH projects are highly dependent on funding. There are funding bodies that provide detailed history of their funded projects and works that can be used to compile new datasets, e.g., UK Research Councils⁷⁸, the National Endowment for the Humanities (NEH) in the US⁷⁹, and the Andrew W. Mellon Foundation⁸⁰.

Journal editorial decision-making is also a crucial factor that could affect the patterns of publication. For example, studies with positive results are more likely to be published than studies with negative results (Olson, 2002). Also, as mentioned in section 4.3, as journal editorial decisions change in light of a more collaborative impulse in publishing, studies have found that co-authored papers have a higher chance of acceptance and higher citation frequencies (Smart and Bayer, 1986). Admittedly, there is some level of publication bias in editorial decision making. Many studies have addressed the issue and explored methods to detect bias and improve the decision making process (Milkman et al., 2009; van Lent et al., 2014; Tennant et al., 2019). It needs to be considered in future studies when constructing datasets.

Structured bibliometric data only presents one perspective of the community. DH is an interdisciplinary field, but its publishing and referencing culture is primarily influenced by the Humanities which is different to that of other disciplines (e.g., Sciences and Engineering) (Becher and Trowler, 2001b). For example, many DH articles use references to support a ‘conversation-like’ discussion in written works, and this is notably common in publications of DH debates and critiques (Berry and Fagerjord, 2017; Gold, 2012; Grusin, 2013; Liu, 2012b). DH scholars use more footnotes than

⁷⁸ The UK Research Councils invest research across all academic disciplines, and in particular, the Arts and Humanities Research Council, the Economic and Social Research Council, and UK Research and Innovation (a cross-council research programme) offers most of the DH-related funding opportunities: <https://ahrc.ukri.org/research/fundedthemesandprogrammes/>

⁷⁹ Projects funded by the National Endowment for the Humanities (NEH): <https://www.neh.gov/our-work>

⁸⁰ The Andrew W. Mellon Foundation’s grants database: <https://mellon.org/grants/grants-database/>

computer scientists (Hellqvist, 2009, p. 311), and many works done by software engineers are not acknowledged in the author byline (Kumar, 2018, pp. 37–38; Poole and Garwood, 2018, p. 184). Collecting citation data without gathering footnotes and acknowledgments and investigating other contributors, will inevitably construct datasets with limited representation (Grafton, 1997). Referencing behaviours also have different sentimental indications that the ACA network is unable to exhibit. Criticism such as negative references are often cited in the publications of humanities topics, while in the natural sciences, it is not common to cite references for negative criticism (Brooks, 1985; Cano, 1989; Meadows, 1974).

Furthermore, bibliometric data has a significant lag time in both citation and co-authorship networks. As mentioned above, in order to build up a citation record for co-citation, it usually needs a lag time of five to eight years (Hopcroft et al., 2004, p. 5250). Given the journal publication process, e.g., the time for peer-review and typesetting, it could be even longer for an idea to be read by sufficient readers to result in citations. This could explain why certain recognisable authors and ‘newcomers’ might not appear on the network yet. Studies have also shown that when a scholar is newly appointed to a post, it usually takes some time before their research ends up in co-authored publications (Rørstad and Aksnes, 2015, p. 319). Additionally, when analysing the bibliometric data from a longitudinal perspective, future studies need to take the annual growth of published papers into consideration. With the advancement of technology and electronic publishing, journals are increasing their publication capacity. Therefore, when studying and comparing bibliometric data between different time periods, the variation in data sample sizes also needs to be noted.

Similarly, the Twitter dataset only presents one part of the informal DH community. Twitter, as a corporate-based platform, has been growing exponentially since its launch in 2006. By March 2010, it had over 70,000 registered users, and this figure grows to over 100 million by 2012, and 330 million monthly active users in 2019 (Molina, 2019). Along with the significant development of the platform, users’ experience and knowledge are maturing and evolving, and the level of interaction and the diversity of discussion topics are also expanding. When collecting data, one should be aware that Twitter, as a data source platform, is under continuous change.

Although the DH community is known to be highly active on social media, especially blogs and Twitter (Ross et al., 2011; Kirschenbaum, 2014), many DH scholars still rarely use Twitter (Galina, 2014; Mahony, 2018) or other social media (Puschmann and Bastos, 2015). For instance, most Chinese DH scholars do not use Twitter, instead, they use WeChat (Mahony, 2018), and most Japanese DH scholars prefer to use Facebook⁸¹ and LINE (Nagasaki and Muller, 2012). The Twitter literacy skill (e.g., reading, interpreting, writing, posting tweets) is seen as a prerequisite to be included in the DH Twitter community, and it requires time and effort to develop (Holmberg and Thelwall, 2014, pp. 1033–1038). There are affordances of the platform itself. For example, there are higher profile verified accounts. Being verified seems to show that a profile is authentic, but it needs to be learned and practiced. Also, the more accounts you follow, the more tweets you get, and the skills such as reordering timelines to focus on ‘top’ rather than ‘latest’ tweets are necessary.

In addition, the representation of geolinguistic diversity may contain unrelated noises. Many non-Anglophone DH scholars choose to use the English-language user interface (this candidate is also one of them), and Grandjean estimated that around 30% of French and German DH users were using Twitter in the English-language interface (Grandjean, 2016, pp. 12–13). Therefore, it is difficult to differentiate them from native English speakers which might lead to a skewed network visualisation.

Additionally, many DH scholars are cautious about using social media and changing traditional scholarly communication patterns, despite the fact that DH is believed to be encouraging the use of digital tools and methods (Weller, 2011). In this study, the scholars extracted from the bibliometric dataset are not the same group of people extracted from the Twitter dataset. As mentioned in section 5.1, there was an evident gap between the DH community in publications and the DH community on Twitter. Despite some level of overlap, this gap may speak to scholarly resistance and critiques towards social media (Sugimoto et al., 2017, p. 2038). For instance, social media was rated as the poorest dissemination method by health policy researchers, describing it as being ‘incompatible with research, of high risk professionally, of uncertain efficacy,

⁸¹ Facebook group of JADH (Japanese Association for Digital Humanities), more information can be found: <https://www.facebook.com/groups/758758500904522/about/>

and an unfamiliar technology that they did not know how to use' (Grande et al., 2014, p. 1278). Twitter, in particular, raises 'negative emotions' when scholars 'critique, connect, and organise', but such emotions, on the other hand, were believed to have the ability to 'fuel the production' (Gregory and Singh, 2018). In addition, networked scholarship such as Twitter exposes the 'vulnerabilities' of scholars and 'creates distinctions and pressures' by making scholars visible and identifiable (Stewart, 2016, pp. 62–63). These concerns may contribute to the gap between the DH bibliometric community and the DH Twitter community, however, there might be other reasons too.

Besides this, we need to keep in mind that there are some scholars who indeed use digital methods to study the humanities and can be recognised as digital humanists, but who do not consider themselves or self-identify as 'digital humanists' (Snyder, 2014).

7.1.2.2 Network visualisation

Network visualisation is a powerful tool for presenting a large number of quantitative elements, and it is a quantified representation, not the actual community (or communities), as Grandjean pointed out in his DH Twitter network visualisation:

These first elements should not make us forget that this network is a visual representation of a set of data whose complexity is not limited to a simple graphical rendering. Beyond a certain aesthetic, sometimes very suggestive, it is in its ability to generate new research questions – pushing the researcher to get back into the data itself – that a network analysis proves his interest. (Grandjean, 2016, p. 4)

Moreover, when interpreting the visualisations, we need to be aware that people are linked and grouped on the network by not one but by a variety of different reasons. This is what Weingart mentioned – 'the Wittgenstein family resemble' – based on a keyword co-occurrence network that he generated from the submissions to DH2013 (Weingart, 2012). According to this concept, nodes that are thought to be connected by one reason may actually be connected by a series of different reasons, where no one reason is common to all of the nodes (Nyström, 2005). For example, nodes on ACA that are thought to be grouped by subject specialties may be connected by audience citing preferences, or one cited author being mentioned in another's work very frequently. Nodes on the densely connected hashtag co-occurrence network are

thought to be grouped by similar semantic meaning or subject relatedness but they may also be grouped by users with a variety of interests or a shorthand term encapsulating different or even opposite meanings (Warwick, 2015, p. 540). Movement participants are also found applying more strategic hashtag combinations to reach different social circles (Wang et al., 2016). Such reasons also potentially explain why semantically related hashtags are sometimes not close or linked on the network (Türker and Sulak, 2017).

Along with the visualisation, although this research has made the effort to investigate the node label display and colour-coding enhancement (not only for aesthetic purposes), there are limitations to be acknowledged. As mentioned in previous chapters, this study tries to avoid identifying individuals, but in order to give examples and provide interpretations of network results, a few scholars and their works are named and discussed. Nevertheless, this study strictly follows UCL's data protection regulations and uses only public information. Information such as gender, location, and language are colour-coded on the network with the possibility for colour blindness taken into consideration (Wong, 2011). However, as some networks contain many clusters (e.g., Figure 5.32 presents 19 colour-coded locations), the potential limitations of such colour encoding method are noted.

7.1.3 Future Study

Although it is not possible to include all the DH scholars that we would like, future studies can expand the dataset (especially with non-Anglophone sources) to demonstrate a more inclusive DH community; as Galina suggested:

What is important to note, however, is that just because DH work is not available in English does not mean that it does not exist, which seems to be the assumption some of these DH studies reach. Any comprehensive study of the DH community would necessarily need to include information available in other languages. (Galina, 2014, p. 310)

For bibliometric analysis, data from other important sources of formal communication, such as a wider range of journals, as well as funding list, conference articles and books, can be collected to improve the DH subject specialty study (Garfield, 1982, p. 5; Hicks and Wang, 2009, pp. 5–11). Non-Anglophone publications, in particular, can be

investigated, such as data sources from Humanistica, RedHD, and the Japanese Association, etc. Apart from data entry for existing fields, more information can be extracted and analysed from the dataset, such as the keywords, abstracts, footnotes, and acknowledgement of each articles.

Similarly, for social media analysis, collecting data from other social networking platforms, such as Facebook, LinkedIn, Yandex, WeChat, LINE, would possibly provide quite different representations of the DH communities (Song et al., 2016, pp. 12–13). Even on Twitter, methods can be developed to discover DH scholars that are not yet connected to ADHO. Although the task would require multi-lingual skills to process the dataset, it might help to construct a more global DH landscape and to understand how DH knowledge is created and shared across different cultures.

As age and professional rank can also be important variables in many fields (Rørstad and Aksnes, 2015), it is worth collecting data from that perspective from both formal and informal channels (e.g., according to publication date and active time period) and exploring the academic and social impact a scholar could have at different stages of academic progression.

Nevertheless, scholars use a variety of channels to communicate and there is no clear boundary between formal and informal types of data. Many data sources contain both, for example, project websites, reading lists, archives, some blogs and conference abstracts. Avoiding such binary division between formal and informal can also help to identify the reasons for the lack of substantial overlap between the current bibliometric dataset and the Twitter dataset.

Apart from expanding the dataset, new methods can be introduced to enhance the results and analysis. From a quantitative aspect, betweenness centrality measures can also be improved in many ways, such as including contributions from all paths between nodes, not just the shortest ones (Newman, 2005) and introducing variants (Brandes, 2008) in order to analyse the network from more aspects. From a qualitative aspect, semi-structured interviews might be conducted to help understand the visualisations. Questions can be asked, e.g., are the DH community and subjects represented here, the ones that would be recognised by the community itself? If the

values of DH on the networks and the ways that they are practiced in real life are different, how would we interpret and make sense out of that?

It is a continuous effort to learn and understand the communities within which we collaborate and carry out research. With the help of the expanded dataset and these new methods, future studies can gain more understandings about 'who we are' and 'what is DH'.

8 Appendix

Appendix A (filename: Appendix A publication_raw_chun_llc_dhq.xlsx) contains publication data collected from *CHum*, *LLC*, *DHQ*, information including metadata of 2,527 articles, and summary of the dataset (e.g., number of articles, references, multi-authored articles).

Appendix B (filename: Appendix B co-author network.xlsx) contains a table of co-authorship network information including number of publications and co-authors, and values of gender, country, and betweenness centrality for each author.

Appendix C (filename: Appendix C co-author international.xlsx) contains tables of international co-authored articles and author information (e.g., if the author has been co-authored internationally with people affiliate in other countries).

Appendix D (filename: Appendix D co-hashtag network.xlsx) contains a table of co-occurred hashtag network information including number of occurrence, betweenness centrality, and topical categories.

Appendix E (filename: Appendix E co-retweet network.xlsx) contains a table of co-retweet network information including number of retweet, values of gender, country, and betweenness centrality for each Twitter user collected.

Appendix F (filename: Appendix F network data.xlsx) contains the map and net files of the four networks that can be opened in VOSviewer, i.e., author co-citation network, co-authorship network, Twitter hashtag co-occurrence network, co-retweet network.

Appendix G (filename: Appendix G high resolution figures.zip) contains the high resolution version of all figures.

9 Bibliography

Abbasi, A., Hossain, L., Leydesdorff, L., 2012. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics* 6, 403–412. <https://doi.org/10.1016/j.joi.2012.01.002>

Abramo, G., D'Angelo, C.A., Caprasecca, A., 2009. Gender differences in research productivity: A bibliometric analysis of the Italian academic system. *Scientometrics* 79, 517–539. <https://doi.org/10.1007/s11192-007-2046-8>

Abramo, G., D'Angelo, C.A., Murgia, G., 2013. Gender differences in research collaboration. *Journal of Informetrics* 7, 811–822. <https://doi.org/10.1016/j.joi.2013.07.002>

Abu Sheikha, F., Inkpen, D., 2010. Automatic classification of documents by formality, in: Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering(NLPKE-2010). Presented at the 2010 International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2010), IEEE, Beijing, pp. 1–5. <https://doi.org/10.1109/NLPKE.2010.5587767>

Acedo, F.J., Barroso, C., Galan, J.L., 2006. The resource-based theory: dissemination and main trends. *Strat. Mgmt. J.* 27, 621–636. <https://doi.org/10.1002/smj.532>

Adams, J., Gurney, K.A., 2018. Bilateral and Multilateral Coauthorship and Citation Impact: Patterns in UK and US International Collaboration. *Front. Res. Metr. Anal.* 3, 12. <https://doi.org/10.3389/frma.2018.00012>

Adema, J., 2014. Open Access, in: Critical Keywords for the Digital Humanities. Meson Press, Lüneburg.

Ahlgren, P., Jarneving, B., Rousseau, R., 2003. Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology* 54, 550–560. <https://doi.org/10.1002/asi.10242>

Ahmed, A., Low, Y., Aly, M., Josifovski, V., Smola, A.J., 2011. Scalable distributed inference of dynamic user interests for behavioral targeting. ACM Press, p. 114. <https://doi.org/10.1145/2020408.2020433>

Ahmed, M., Gil, A., Pereira, M.P. de S., Risam, R., Álvarez, M.E., Fernández, S.A., Rodriguez, L., Martinez, M., 2018. Editors' Choice: Torn Apart / Separados [WWW Document]. Digital Humanities Now. URL <http://digitalhumanitiesnow.org/2018/06/editors-choice-torn-apart-separados/> (accessed 12.26.19).

Aksnes, D.W., Rorstad, K., Piro, F., Sivertsen, G., 2011. Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology* 62, 628–636. <https://doi.org/10.1002/asi.21486>

Aldrich, H.E., 2012. The emergence of entrepreneurship as an academic field: A personal essay on institutional entrepreneurship. *Research Policy* 41, 1240–1248. <https://doi.org/10.1016/j.respol.2012.03.013>

Alexa, 2013. Alexa Top 500 Global Sites [WWW Document]. Alexa, an Amazon company. URL <https://www.alexa.com/topsites> (accessed 10.25.18).

Algarni, M.A., 2014. The Use of Social Media In Informal Scientific Communication among Scholars: Modeling The Modern Invisible College (Doctor of philosophy). University of North Texas.

Alonso, J.M., Castiello, C., Mencar, C., 2018. A Bibliometric Analysis of the Explainable Artificial Intelligence Research Field, in: Medina, J., Ojeda-Aciego, M., Verdegay, J.L., Pelta, D.A., Cabrera, I.P., Bouchon-Meunier, B., Yager, R.R. (Eds.), *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*. Springer International Publishing, Cham, pp. 3–15. https://doi.org/10.1007/978-3-319-91473-2_1

Alvarado, R.C., 2012. The Digital Humanities Situation, in: *Debates in the Digital Humanities*. p. 50.

Ammon, U., 2012. Linguistic inequality and its effects on participation in scientific discourse and on global knowledge accumulation – With a closer look at the problems of the second-rank language communities. *Applied Linguistics Review* 3. <https://doi.org/10.1515/applrev-2012-0016>

Ammon, U., 2006. Language Planning for International Scientific Communication: An Overview of Questions and Potential Solutions. *Current Issues in Language Planning* 7, 1–30. <https://doi.org/10.2167/cilp088.0>

Anheier, H.K., Gerhards, J., Romo, F.P., 1995. Forms of Capital and Social Structure in Cultural Fields: Examining Bourdieu's Social Topography. *American Journal of Sociology* 100, 859–903. <https://doi.org/10.1086/230603>

Archambault, A., Grudin, J., 2012. A longitudinal study of facebook, linkedin, and twitter use, in: *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*. Presented at the the 2012 ACM annual conference, ACM Press, Austin, Texas, USA, p. 2741. <https://doi.org/10.1145/2207676.2208671>

Artwick, C.G., 2014. News sourcing and gender on Twitter. *Journalism* 15, 1111–1127. <https://doi.org/10.1177/1464884913505030>

Backhaus, K., Lügger, K., Koch, M., 2011. The structure and evolution of business-to-business marketing: A citation and co-citation analysis. *Industrial Marketing Management* 40, 940–951. <https://doi.org/10.1016/j.indmarman.2011.06.024>

Badar, K., Hite, J.M., Badir, Y.F., 2013. Examining the relationship of co-authorship network centrality and gender on academic research performance: the case of chemistry researchers in Pakistan. *Scientometrics* 94, 755–775. <https://doi.org/10.1007/s11192-012-0764-z>

Bailey, M., Cong-Huyen, A., Lothian, A., Phillips, A., 2016. *Reflections on a Movement: #transformDH, Growing Up*, in: *Debates in the Digital Humanities: 2016*. University of Minnesota Press, Minneapolis London.

Bailey, M.Z., 2011. All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave. *Journal of Digital Humanities* 1.

Bailey, R.W., Gorlach, M., Arbor, A., 1986. English as a World Language. *RELC Journal* 17, 91–96. <https://doi.org/10.1177/003368828601700107>

Baker, P., Hardie, A., McEnery, T., Xiao, R., Bontcheva, K., Cunningham, H., Gaizauskas, R., Hamza, O., Maynard, D., Tablan, V., Ursu, C., Jayaram, B.D., Leisher, M., 2004. Corpus Linguistics and South Asian Languages: Corpus Creation and Tool Development. *Lit Linguist Computing* 19, 509–524. <https://doi.org/10.1093/linc/19.4.509>

Barjak, F., 2006. Research productivity in the internet era. *Scientometrics* 68, 343–360. <https://doi.org/10.1007/s11192-006-0116-y>

Barnett, F., 2014. The Brave Side of Digital Humanities. *differences* 25, 64–78. <https://doi.org/10.1215/10407391-2420003>

Bartko, J.J., 1966. The Intraclass Correlation Coefficient as a Measure of Reliability. *Psychol Rep* 19, 3–11. <https://doi.org/10.2466/pr0.1966.19.1.3>

Basow, S.A., Rubenfeld, K., 2003. “Troubles Talk”: Effects of Gender and Gender-Typing. *Sex Roles* 48, 183–187. <https://doi.org/10.1023/A:1022411623948>

BBC, 2013. Bomb threat tweet sent to Mary Beard. BBC News.

Becher, T., Trowler, P., 2001a. Academic tribes and territories . Buckingham: Society for Research in Higher Education. Open University Press.

Becher, T., Trowler, P., 2001b. Academic tribes and territories: intellectual enquiry and the culture of disciplines, 2nd ed. ed. Open University Press, Philadelphia, PA.

Bennett, K., 2011. Academic writing in Portugal. Imprensa da Universidade de Coimbra, Coimbra.

Berry, D.M., 2011. The philosophy of software code and mediation in the digital age. Palgrave Macmillan, Basingstoke; New York.

Berry, D.M., Fagerjord, A. (Eds.), 2017. Digital humanities: knowledge and critique in a digital age. Malden, MA : Polity, Cambridge, England.

Bianco, J., 2012. The digital humanities which is not one, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. U of Minnesota Press, pp. 96–112.

Biber, D., 1988. Variation across speech and writing, transferred to digit. pr. ed. Cambridge Univ. Press, Cambridge.

Blashfield, R.K., Reynolds, S.M., 2012. An Invisible College View of the DSM-5 Personality Disorder Classification. *Journal of Personality Disorders* 26, 821–829.

Blumell, L.E., 2019. She Persisted ... and so did He: Gendered source use during the Trump Access Hollywood scandal. *Journalism Studies* 20, 267–286. <https://doi.org/10.1080/1461670X.2017.1360150>

Bocanegra-Valle, A., 2014. ‘English is my default academic language’: Voices from LSP scholars publishing in a multilingual journal. *Journal of English for Academic Purposes* 13, 65–77. <https://doi.org/10.1016/j.jeap.2013.10.010>

Bolter, J.D., Grusin, R., 2003. Remediation: understanding new media, 6. Nachdr. ed. MIT Press, Cambridge, Mass.

Bond, S.E., Long, H., Underwood, T., 2017. ‘Digital’ Is Not the Opposite of ‘Humanities.’ *The Chronicle of Higher Education*.

Boonstra, O., Breure, L., Doorn, P., 2006. Past, present and future of historical information science. DANS. https://doi.org/10.26530/OAPEN_353255

Borg, I., Groenen, P.J.F., 2005. Modern Multidimensional Scaling, Springer Series in Statistics. Springer New York, New York, NY.

Borgman, C., 2009. Scholarship in the Digital Age: Information, Infrastructure, and the Internet. Columbia University, New York.

Bork, P., Jensen, L.J., von Mering, C., Ramani, A.K., Lee, I., Marcotte, E.M., 2004. Protein interaction networks from yeast to human. *Current Opinion in Structural Biology* 14, 292–299. <https://doi.org/10.1016/j.sbi.2004.05.003>

Börner, K., 2011. Plug-and-play macroscopes. *Communications of the ACM* 54, 60. <https://doi.org/10.1145/1897852.1897871>

Börner, K., Chen, C., Boyack, K.W., 2005. Visualizing knowledge domains. *Ann. Rev. Info. Sci. Tech.* 37, 179–255. <https://doi.org/10.1002/aris.1440370106>

Bowker, L., 2018. Corpus linguistics is not just for linguists: Considering the potential of computer-based corpus methods for library and information science research. *Library Hi Tech* 36, 358–371. <https://doi.org/10.1108/LHT-12-2017-0271>

Bowman, T.D., Demarest, B., Weingart, S.B., Simpson, G.L., Larivière, V., Thelwall, M., Sugimoto, C.R., 2013. Mapping DH through heterogeneous communicative practices [WWW Document]. Indiana University Bloomington. URL <http://did.ils.indiana.edu/dh/pdf/DH2013.MappingDH.pdf>

Boyd, D., Crawford, K., 2012. Critical Questions For Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15, 662–679. <https://doi.org/10.1080/1369118X.2012.678878>

Boyd, D., Golder, S., Lotan, G., 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. IEEE, pp. 1–10. <https://doi.org/10.1109/HICSS.2010.412>

Bozeman, B., Gaughan, M., 2011. How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy* 40, 1393–1402. <https://doi.org/10.1016/j.respol.2011.07.002>

Bradley, J., 2016. No Job for Techies: Technical Contributions to Research in the Digital Humanities, in: Mccarty, W., Deegan, M. (Eds.), *Collaborative Research in the Digital Humanities*. Routledge, London, UK, pp. 23–38. <https://doi.org/10.4324/9781315572659-6>

Brandeis Library, 2013. » Digital Humanities: From 1851? Brandeis Library. Brandeis Library. URL <https://blogs.brandeis.edu/library/2013/05/17/digital-humanities-from-1851/> (accessed 4.5.19).

Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* 30, 136–145. <https://doi.org/10.1016/j.socnet.2007.11.001>

Brandes, U., 2001. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology* 25, 163–177. <https://doi.org/10.1080/0022250X.2001.9990249>

Brass, D.J., 1985. Men's and Women's Networks: A Study of Interaction Patterns and Influence In an Organization. *Academy of Management Journal* 28, 327–343. <https://doi.org/10.5465/256204>

Bray, F., 2007. Gender and Technology. *Annu. Rev. Anthropol.* 36, 37–53. <https://doi.org/10.1146/annurev.anthro.36.081406.094328>

Brennan, C., 2018. Digital humanities, digital methods, digital history, and digital outputs: History writing and the digital revolution. *History Compass* 16, e12492. <https://doi.org/10.1111/hic3.12492>

Brennan, S., 2015. DH Centered in Museums? [WWW Document]. Lot 49. URL <http://www.lotfortynine.org/2015/03/dh-centered-in-museums/> (accessed 5.14.19).

Brockman, W.S., Neumann, L., Palmer, C.L., Tidline, T.J. (Eds.), 2001. Scholarly work in the humanities and the evolving information environment, Optimizing collections and services for scholarly use. Digital Library Federation, Council on Library and Information Resources, Washington, D.C.

Brooks, T.A., 1985. Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science* 36, 223–229. <https://doi.org/10.1002/asi.4630360402>

Bruns, A., 2012. How Long is a Tweet? Mapping Dynamic Conversation Networks on Twitter Using Gawk and Gephi. *Information, Communication & Society* 15, 1323–1351. <https://doi.org/10.1080/1369118X.2011.635214>

Bruns, A., Weller, K., 2016. Twitter as a first draft of the present: and the challenges of preserving it for the future, in: *Proceedings of the 8th ACM Conference on Web Science - WebSci '16*. Presented at the the 8th ACM Conference, ACM Press, Hannover, Germany, pp. 183–189. <https://doi.org/10.1145/2908131.2908174>

Bryant, K., 2011. Black Girls Code [WWW Document]. imagine. build. create. - Black Girls Code, BlackGirlsCode, Women of Color in Technology. URL <http://www.blackgirlscode.com/> (accessed 1.23.19).

Bryson, T., Posner, M., St. Pierre, A., Varner, S., 2011. SPEC Kit 326: Digital Humanities (November 2011).

Bu, Y., Liu, T., Huang, W., 2016. MACA: a modified author co-citation analysis method combined with general descriptive metadata of citations. *Scientometrics* 108, 143–166. <https://doi.org/10.1007/s11192-016-1959-5>

Bullas, J., 2013. 9 Simple and Powerful Ways to Get More Retweets on Twitter: Report. Jeffbullas's Blog. URL <https://www.jeffbullas.com/9-simple-and-powerful-ways-to-get-more-retweets-on-twitter-report/> (accessed 3.20.19).

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., Schnapp, J., 2012. *Digital_Humanities*. MIT Press.

Burgoyne, J.A., Fujinaga, I., Downie, J.S., 2015. Music Information Retrieval, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A New Companion to Digital Humanities*. John Wiley & Sons, Ltd, Chichester, UK, pp. 213–228. <https://doi.org/10.1002/9781118680605.ch15>

Burrows, J., 2002. "Delta": a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing* 17, 267–287. <https://doi.org/10.1093/llc/17.3.267>

Burrows, J.F., Craig, D.H., 1994. Lyrical drama and the "turbid mountebanks": Styles of dialogue in romantic and renaissance tragedy. *Computers and the Humanities* 28, 63–86. <https://doi.org/10.1007/BF01830688>

Burt, R.S., 2005. Brokerage and closure: an introduction to social capital, Clarendon lectures in management studies series. Oxford University Press, Oxford ; New York.

Burton, D.M., 1982. Automated concordances and word-indexes: Machine decisions and editorial revisions. *Computers and the Humanities* 16, 195–218. <https://doi.org/10.1007/BF02263544>

Burton, D.M., 1981a. Automated concordances and word indexes: The fifties. *Computers and the Humanities* 15, 1–14. <https://doi.org/10.1007/BF02404370>

Burton, D.M., 1981b. Automated concordances and word indexes: The early sixties and the early centers. *Computers and the Humanities* 15, 83–100. <https://doi.org/10.1007/BF02404202>

Burton, D.M., 1981c. Automated concordances and word indexes: The process, the programs, and the products. *Computers and the Humanities* 15, 139–154. <https://doi.org/10.1007/BF02404180>

Burton, M., 2015. Blogs as Infrastructure for Scholarly Communication (PhD Thesis). University of Michigan, Ann Arbor, Michigan, USA.

Busa, R., 2004. Foreword: Perspectives on the Digital Humanities, in: A Companion to Digital Humanities. Blackwell Publishing, Malden, MA/Oxford/Carlton, Victoria, pp. xvi–xxi.

Busa, R., 1980. The annals of humanities computing: The index Thomisticus. *Computers and the Humanities* 14, 83–90. <https://doi.org/10.1007/BF02403798>

Busa, R., 1950. Complete Index Verborum of Works of St. Thomas. *Speculum: A Journal of Medieval Studies* XXV, 424–425.

Busa, R., 1949. La terminologia Tomistica dell'Interiorità: Saggi di metodo per un'interpretazione della metafisica della presenza. Fratelli Bocca, Milano.

Busa, R., 1948. A proposito di immanenza e transcendenza. Presented at the Congresso Internazionale de Filosofia, Madrid, pp. 671–689.

Buurma, R.S., Heffernan, L., 2018. Search and Replace: Josephine Miles and the Origins of Distant Reading [WWW Document]. Modernism / Modernity. URL <https://modernismmodernity.org/forums/posts/search-and-replace> (accessed 7.1.19).

Cano, V., 1989. Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science* 40, 284–290. [https://doi.org/10.1002/\(SICI\)1097-4571\(198907\)40:4<284::AID-ASI10>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(198907)40:4<284::AID-ASI10>3.0.CO;2-Z)

Carrington, P.J., Scott, J., Wasserman, S., 2005. Models and Methods in Social Network Analysis. Cambridge University Press.

Carrotte, E.R., Prichard, I., Lim, M.S.C., 2017. "Fitspiration" on Social Media: A Content Analysis of Gendered Images. *J Med Internet Res* 19, e95. <https://doi.org/10.2196/jmir.6368>

Casanueva, C., Larrinaga, C., 2013. The (uncertain) invisible college of Spanish accounting scholars. *Critical Perspectives on Accounting* 24, 19–31. <https://doi.org/10.1016/j.cpa.2012.05.002>

Cecire, N., 2012. Introduction: Theory and the Virtues of Digital Humanities. *Journal of Digital Humanities*.

Cecire, N., 2011. When Digital Humanities Was in Vogue. *Journal of Digital Humanities* 1.

centerNet, 2018. An international network of digital humanities centers [WWW Document]. CenterNet. URL <https://dhcenternet.org/> (accessed 10.9.16).

Chan, T., Joshi, N., Lin, M., Mehta, N., 2015. Using Google Hangouts on Air for Medical Education: A Disruptive Way to Leverage and Facilitate Remote Communication and Collaboration. *Journal of Graduate Medical Education* 7, 171–173. <https://doi.org/10.4300/JGME-D-14-00545.1>

Chang Boon Lee, P., 2003. Going beyond career plateau: Using professional plateau to account for work outcomes. *Journal of Mgmt Development* 22, 538–551. <https://doi.org/10.1108/02621710310478503>

Chang, T.-K., Tai, Z., 2005. Mass Communication Research and the Invisible College Revisited: The Changing Landscape and Emerging Fronts in Journalism-Related Studies. *Journalism & Mass Communication Quarterly* 82, 672–694. <https://doi.org/10.1177/107769900508200312>

Chen, K., Hsueh, B.-S., 2013. Exploring Regional Development of Digital Humanities Research: A Case Study for Taiwan. *arXiv:1312.5817 [cs]*.

Cheng, F.-F., Huang, Y.-W., Tsaih, D.-C., Wu, C.-S., 2019. Trend analysis of co-authorship network in Library Hi Tech. *Library Hi Tech* 37, 43–56. <https://doi.org/10.1108/LHT-11-2017-0241>

Chodorow, N., 1999. The reproduction of mothering: psychoanalysis and the sociology of gender: with a new preface. University of California Press, Berkeley.

Chomsky, N., 1965. *Aspects of the theory of syntax*, 19. pr. ed, Special technical report of the Research Laboratory of Electronics of the Massachusetts Institute of Technology. MIT Press, Cambridge/Mass.

Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-Law Distributions in Empirical Data. *SIAM Rev.* 51, 661–703. <https://doi.org/10.1137/070710111>

Clavert, F., 2012. The Digital Humanities multicultural revolution did not happen yet. L'histoire contemporaine à l'ère numérique. URL <https://histnum.hypotheses.org/1546> (accessed 10.13.18).

CNKI, 2019. CNKI website [WWW Document]. report. URL <http://oversea.cnki.net/kns55/support/en/products.aspx> (accessed 4.15.19).

Coffee, N., Koenig, J.-P., Poornima, S., Forstall, C.W., Ossewaarde, R., Jacobson, S.L., 2013. The Tesserae Project: intertextual analysis of Latin poetry. *Literary and Linguistic Computing* 28, 221–228. <https://doi.org/10.1093/linc/fqs033>

Cohen, D., 2019. digitalhumanities - list members [WWW Document]. Twitter. URL <https://twitter.com/dancohen/lists/digitalhumanities/members?lang=en> (accessed 3.11.19).

Cohen, D., 2009. Introducing Digital Humanities Now – Dan Cohen. URL <https://dancohen.org/2009/11/18/introducing-digital-humanities-now/> (accessed 3.11.19).

Cole, K.K., 2015. “It’s Like She’s Eager to be Verbally Abused”: Twitter, Trolls, and (En)Gendering Disciplinary Rhetoric. *Feminist Media Studies* 15, 356–358. <https://doi.org/10.1080/14680777.2015.1008750>

Collins, K., Shiffman, D., Rock, J., 2016. How Are Scientists Using Social Media in the Workplace? *PLOS ONE* 11, e0162680. <https://doi.org/10.1371/journal.pone.0162680>

Côté, I.M., Darling, E.S., 2018. Scientists on Twitter: Preaching to the choir or singing from the rooftops? *FACETS* 3, 682–694. <https://doi.org/10.1139/facets-2018-0002>

Crane, D., 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago, University of Chicago Press.

Cronin, B., 2005. *The hand of science: academic writing and its rewards*. Scarecrow Press, Lanham, Md.

Cronin, B., Sugimoto, C.R., 2014. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. MIT Press.

Crymble, A., 2016. Digital library search preferences amongst historians and genealogists: British History Online user survey. *DHQ* 10.

Curtis, S., 2013. Teaching our children to code: a quiet revolution.

Dacos, M., 2019. digitalhumanities - list members [WWW Document]. Twitter. URL <https://twitter.com/marindacos/lists/digitalhumanities/members> (accessed 3.11.19).

Dacos, M., 2013. La estrategia de la sauna finlandesa. *Humanidades Digitales*. URL <http://humanidadesdigitales.net/blog/2013/08/12/estrategia-sauna-finlandesa/>

Dacos, M., 2012. results of the online survey “Who are you, Digital Humanists?”

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>

De la Cruz, D.M., Kaupp, J., Kemman, M., Lewis, K., 2015. Mapping Cultures in the Big Tent: Multidisciplinary Networks in the Digital Humanities Quarterly, in: IVMOOC 2015. Presented at the IVMOOC 2015 - Visualizing the Digital Humanities Project.

De Solla Price, D.J., 1986. *Little Science, Big Science ... and beyond*. Columbia University Press, New York.

De Solla Price, D.J., 1963. *Little Science, Big Science*. Columbia University Press, New York.

De Solla Price, D.J., Beaver, D., 1966. Collaboration in An Invisible College. *American Psychologist* 21, 1011–1018. <https://doi.org/10.1037/h0024051>

De Stefano, D., Giordano, G., Vitale, M.P., 2011. Issues in the analysis of co-authorship networks. *Quality & Quantity* 45, 1091–1107. <https://doi.org/10.1007/s11135-011-9493-2>

Deegan, M., McCarty, W., 2012. *Collaborative Research in the Digital Humanities*. Ashgate Publishing, Ltd.

DH at Berkeley, 2019. digitalhumanities - list members [WWW Document]. Twitter. URL <https://twitter.com/DHBerkeley/lists/digital-at-berkeley> (accessed 3.11.19).

DHQ, 2017. We start 2017 with some great news - @DHQuarterly is now indexed in Thomson Reuters' Web of Science, in the Emerging Sources Citation Index! @DHQuarterly. URL <https://twitter.com/DHQuarterly/status/817046447036268544> (accessed 2.14.19).

Digital Scholarship Group, 2012. Digital Humanities Quarterly [WWW Document]. Northeastern University. URL <https://dsg.neu.edu/projects/dhq/> (accessed 1.9.19).

Ding, Y., 2011. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics* 5, 187–203. <https://doi.org/10.1016/j.joi.2010.10.008>

Dinsman, M., 2016. The Digital in the Humanities: An Interview with Franco Moretti [WWW Document]. Los Angeles Review of Books. URL <https://lareviewofbooks.org/article/the-digital-in-the-humanities-an-interview-with-franco-moretti/> (accessed 10.4.18).

DLF, 2020. Digital Library Federation [WWW Document]. DLF. URL <https://www.diglib.org/> (accessed 4.4.20).

D'Monte, L., 2009. Swine flu's tweet tweet causes online flutter. Business Standard India.

Doctor, V., 2012. What Characters Can A Hashtag Include? [WWW Document]. hashtags.org. URL <https://www.hashtags.org/featured/what-characters-can-a-hashtag-include/> (accessed 12.28.19).

Duszak, A., Lewkowicz, J., 2008. Publishing academic texts in English: A Polish perspective. *Journal of English for Academic Purposes* 7, 108–120. <https://doi.org/10.1016/j.jeap.2008.03.001>

Eagly, A.H., 2013. *Sex Differences in Social Behavior: A Social-role interpretation*, 1st ed. Psychology Press. <https://doi.org/10.4324/9780203781906>

Earhart, A.E., 2018. Digital Humanities Within a Global Context: Creating Borderlands of Localized Expression. *Fudan Journal of the Humanities and Social Sciences* 11, 357–369. <https://doi.org/10.1007/s40647-018-0224-0>

Earhart, A.E., 2012. Can Information Be Unfettered?: Race and the New Digital Humanities Canon, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. University of Minnesota Press, Minneapolis, pp. 309–318.

Earhart, A.E., Risam, R., Bruno, M., 2020. Gendered Citational Politics: The Challenges of Quantifying Impact in Digital Scholarship in the Humanities. *Digital Scholarship in the Humanities*.

Eck, N.J. van, Waltman, L., 2009. How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology* 60, 1635–1651. <https://doi.org/10.1002/asi.21075>

Ediger, D., Jiang, K., Riedy, J., Bader, D.A., Corley, C., 2010. Massive Social Network Analysis: Mining Twitter for Social Good, in: 2010 39th International Conference on Parallel Processing. Presented at the 2010 39th International Conference on Parallel Processing (ICPP), IEEE, San Diego, CA, USA, pp. 583–593. <https://doi.org/10.1109/ICPP.2010.66>

Edwards, C., 2012. The digital humanities and its users, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. U of Minnesota Press, pp. 213–232.

Eiteljorg, H., 2004. Computing for Archaeologists, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A Companion to Digital Humanities*. Blackwell Publishing Ltd, Malden, MA, USA, pp. 20–30. <https://doi.org/10.1002/9780470999875.ch2>

Elteren, M. van, 2003. U.S. Cultural Imperialism: Today Only a Chimera. *SAIS Review* 23, 169–188. <https://doi.org/10.1353/sais.2003.0038>

EMiC, 2010. Home [WWW Document]. Editing Modernism in Canada. URL <http://editingmodernism.ca/> (accessed 4.27.20).

Eom, S.B., 2003. *Author Co-citation Analysis Using Custom Bibliographic Databases: An Introduction to the SAS Approach*. Edwin Mellen Press.

Erford, B.T., Miller, E.M., Schein, H., McDonald, A., Ludwig, L., Leishear, K., 2011. Publication Patterns: Author and Article Characteristics From 1994 to 2009. *Journal of Counseling & Development* 89, 73–80. <https://doi.org/10.1002/j.1556-6678.2011.tb00062.x>

Eriksson Krutrök, M., Lindgren, S., 2018. Continued Contexts of Terror: Analyzing Temporal Patterns of Hashtag Co-Occurrence as Discursive Articulations. *Social Media + Society* 4, 205630511881364. <https://doi.org/10.1177/2056305118813649>

Eurobarometer, 2012. Eurobarometer report – Europeans and their languages, June 2012 [WWW Document]. Eurobarometer. URL https://web.archive.org/web/20160106183351/http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf (accessed 2.25.20).

Eysenbach, G., 2011. Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research* 13, e123. <https://doi.org/10.2196/jmir.2012>

Fagan, J.M., Eddens, K., Dolly, J., Vanderford, N., Weiss, H., Levens, J., 2018. Assessing Research Collaboration through Co-authorship Network Analysis.

Fatima, M., Hasan, K., Anwar, S., Nawab, R.M.A., 2017. Multilingual author profiling on Facebook. *Information Processing & Management* 53, 886–904. <https://doi.org/10.1016/j.ipm.2017.03.005>

Favero, M.D., 2010. Academic Disciplines - Disciplines and the Structure of Higher Education, Discipline Classification Systems, Discipline Differences [WWW Document]. State University. URL

<http://education.stateuniversity.com/pages/1723/Academic-Disciplines.html>
(accessed 4.3.16).

Ferguson, G., Pérez-Llantada, C., Plo, R., 2011. English as an international language of scientific publication: a study of attitudes: English as an international language of scientific publication. *World Englishes* 30, 41–59. <https://doi.org/10.1111/j.1467-971X.2010.01656.x>

Fiornonte, D., 2017. Digital Humanities and the Geopolitics of Knowledge. *Digital Studies/Le champ numérique* 7, 5. <https://doi.org/10.16995/dscn.274>

Fiornonte, D., 2014. Digital Humanities from a global perspective. *Laboratorio dell'ISPF*. <https://doi.org/10.12862/ispf14L203>

Fiornonte, D., 2012. Towards a Cultural Critique of the Digital Humanities. *Hist. Soc. Res.* 37, 59–76.

Fitzgerald, F.S., Wilson, E., 2009. The crack-up: with other miscellaneous pieces, excerpts from note-books and letters by F. Scott Fitzgerald together with letters to Fitzgerald from Gertrude Stein, Edith Wharton, T.S. Eliot, Thomas Wolfe, and John Dos Passos, and essays and poems by Paul Rosenfeld, Glenway Wescott, John Dos Passos, John Peale Bishop, and Edmj d Wilson.

Fitzpatrick, K., 2011. *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York.

Fitzpatrick, K., 2010. The Stakes of Disciplinarity [WWW Document]. *Planned Obsolescence*. URL <https://web.archive.org/web/20170302142726/http://www.plannedobsolescence.net/the-stakes-of-disciplinarity/> (accessed 6.26.19).

Flanders, J., 2019. About DHQ [WWW Document]. *Digital Humanities Quarterly (DHQ)*. URL <http://www.digitalhumanities.org/dhq/about/about.html> (accessed 1.9.19).

Flanders, J., 2016. Collaboration and Dissent: Challenges of Collaborative Standards for Digital Humanities, in: *Collaborative Research in the Digital Humanities*. Ashgate, Surrey, pp. 67–80. <https://doi.org/10.4324/9781315572659-8>

Flanders, J., Piez, W., Terras, M., 2007. Welcome to Digital Humanities Quarterly 1.

Flowerdew, J., 2000. Discourse Community, Legitimate Peripheral Participation, and the Nonnative-English-Speaking Scholar. *TESOL Quarterly* 34, 127. <https://doi.org/10.2307/3588099>

Flowerdew, J., 1999. Writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing* 8, 123–145. [https://doi.org/10.1016/S1060-3743\(99\)80125-8](https://doi.org/10.1016/S1060-3743(99)80125-8)

Fluharty, S., 2010. Gender Breakdown for Followers of Digital Humanities on Twitter [WWW Document]. Google Groups. URL <https://groups.google.com/forum/#topic/digital-humanities-barriers-to-participation/wXGHa-UqhRE> (accessed 10.25.18).

Fong, W., Watt, J.C.Y. (Eds.), 1996. *Possessing the past: treasures from the National Palace Museum, Taipei*. Metropolitan Museum of Art: Distributed by H.N. Abrams, New York.

Forte, M., 2015. Cyber Archaeology: A Post-virtual Perspective, in: Svensson, P., Goldberg, D.T. (Eds.), *Between Humanities and the Digital*. MITP.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. *Social Networks* 1, 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)

Freeman, L.C., 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry* 40, 35. <https://doi.org/10.2307/3033543>

French, A., 2009. Make “10” louder, or, the amplification of scholarly communication. amandafrench.net. URL <http://amandafrench.net/2009/12/30/make-10-louder/> (accessed 10.25.18).

Frenken, K., Ponds, R., Van Oort, F., 2010. The citation impact of research collaboration in science-based industries: A spatial-institutional analysis: The citation impact of research collaboration. *Papers in Regional Science* 89, 351–271. <https://doi.org/10.1111/j.1435-5957.2010.00309.x>

Fujinaga, I., Weiss, S.F., 2004. Music, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A Companion to Digital Humanities*. Blackwell Publishing Ltd, Malden, MA, USA, pp. 97–107. <https://doi.org/10.1002/9780470999875.ch9>

Galina, I., 2014. Geographical and linguistic diversity in the Digital Humanities. *Literary and Linguistic Computing* 29, 307–316. <https://doi.org/10.1093/lrc/fqu005>

Galina, I., 2013. Is there anybody out there? building a global digital humanities community. Humanidades Digitales. URL <http://humanidadesdigitales.net/blog/2013/07/19/is-there-anybody-out-there-building-a-global-digital-humanities-community/> (accessed 1.31.18).

Gallon, K., 2016. Making a Case for the Black Digital Humanities, in: *Debates in the Digital Humanities: 2016*. University of Minnesota Press, Minneapolis London.

Gardiner, E., Musto, R.G., 2015. *The Digital Humanities: A Primer for Students and Scholars*. Cambridge University Press, Cambridge.

Garfield, E., 1982. Data from Arts & Humanities Citation Index reveal the interrelationships of science and humanities. *Current Contents* 46, 5–7.

Garfield, E., 1979. Is citation analysis a legitimate evaluation tool? *Scientometrics* 1, 359–375. <https://doi.org/10.1007/BF02019306>

Garfield, Eugene, 1979. Citation indexing - its theory and application in science, technology, and humanities, *Information sciences series*. Wiley, New York.

Garfield, E., 1976. English – An International language for science. *The Information Scientist* 1967, 19–20.

Garfield, E., 1955. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science* 122, 108–111. <https://doi.org/10.1126/science.122.3159.108>

Garvey, W.D., 1979. Communication Facilitating Information Exchange Among Librarians, Scientists, Engineers and Students. Elsevier Science, Kent.

Gascoigne, T., Cheng, D., Claessens, M., Metcalfe, J., Schiele, B., Shi, S., 2010. Is science communication its own field? *JCOM: Journal of Science Communication* 9.

Gazni, A., Sugimoto, C.R., Didegah, F., 2012. Mapping world scientific collaboration: Authors, institutions, and countries. *J. Am. Soc. Inf. Sci.* 63, 323–335. <https://doi.org/10.1002/asi.21688>

Gelley, B., John, A., 2015. Do I Need To Follow You?: Examining the Utility of The Pinterest Follow Mechanism. ACM Press, pp. 1751–1762. <https://doi.org/10.1145/2675133.2675209>

Georghiou, L., 1998. Global cooperation in research. *Research Policy* 27, 611–626. [https://doi.org/10.1016/S0048-7333\(98\)00054-7](https://doi.org/10.1016/S0048-7333(98)00054-7)

Giachanou, A., Crestani, F., 2016. Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Comput. Surv.* 49, 1–41. <https://doi.org/10.1145/2938640>

Gibbs, F., 2013. Digital Humanities Definitions by Type, in: Terras, M., Nyhan, J., Vanhoutte, E. (Eds.), *Defining Digital Humanities: A Reader*. Routledge, London.

Gil, A., Ortega, É., 2016. Global outlooks in digital humanities: multilingual practices and minimal computing, in: Crompton, C., Lane, R.J., Siemens, R. (Eds.), *Doing Digital Humanities: Practice, Training, Research*. Taylor & Francis Group, London, UK.

Gilligan, C., 1993. *In a different voice: psychological theory and women's development*. Harvard University Press, Cambridge, Mass.

Gladwell, M., 2008. *Outliers: the story of success*, 1st ed. ed. Little, Brown and Co, New York.

Glänzel, W., 2001. National characteristics in international scientific co-authorship relations. *Scientometrics* 51, 69–115. <https://doi.org/10.1023/A:1010512628145>

Glänzel, W., Schubert, A., 2005a. Analysing Scientific Networks Through Co-Authorship, in: Moed, H.F., Glänzel, W., Schmoch, U. (Eds.), *Handbook of Quantitative Science and Technology Research*. Kluwer Academic Publishers, Dordrecht, pp. 257–276. https://doi.org/10.1007/1-4020-2755-9_12

Glänzel, W., Schubert, A., 2005b. Domesticity and internationality in co-authorship, references and citations. *Scientometrics* 65, 323–342. <https://doi.org/10.1007/s11192-005-0277-0>

Glänzel, W., Schubert, A., 2001. Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics* 50, 199–214. <https://doi.org/10.1023/A:1010561321723>

Gmür, M., 2003. Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics* 57, 27–57. <https://doi.org/10.1023/A:1023619503005>

Gold, M.K., 2012. *Debates in the Digital Humanities*. U of Minnesota Press.

Gold, M.K., Klein, L.F. (Eds.), 2016. *Debates in the digital humanities: 2016*. University of Minnesota Press, Minneapolis London.

González-Blanco, E., Cantón, C.M., del Rio Riande, G., Ros, S., Pastor, R., Robles-Gómez, A., Caminero, A., Platas, M.L.D., del Olmo, Á., Urízar, M., 2017. EVI-

LINHD, a virtual research environment for the Spanish-speaking community. *Digital Scholarship Humanities* 32, ii171–ii178. <https://doi.org/10.1093/llc/fqx025>

Gonzalez-Brambila, C., Veloso, F.M., 2007. The determinants of research output and impact: A study of Mexican researchers. *Research Policy* 36, 1035–1051. <https://doi.org/10.1016/j.respol.2007.03.005>

Goodman, A.C., Goodman, J., Goodman, L., Goodman, S., 2015. A Few Goodmen: Surname-sharing Economist Coauthors. *Economic Inquiry* 53, 1392–1395. <https://doi.org/10.1111/ecin.12167>

Goodman, S.W., 2011. Controlling Immigration through Language and Country Knowledge Requirements. *West European Politics* 34, 235–255. <https://doi.org/10.1080/01402382.2011.546569>

Google, 2018. Google Arts & Culture [WWW Document]. URL <https://artsandculture.google.com/> (accessed 11.17.18).

Gordon, M.D., 1980. A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. *Scientometrics* 2, 193–201. <https://doi.org/10.1007/BF02016697>

Goswami, S., Sarkar, S., Rustagi, M., 2009. Stylometric Analysis of Bloggers' Age and Gender, in: ICWSM. Presented at the the Third International ICWSM Conference (2009).

Grafton, A., 1997. The footnote: a curious history, Rev. ed. ed. Harvard University Press, Cambridge, Mass.

Grande, D., Gollust, S.E., Pany, M., Seymour, J., Goss, A., Kilaru, A., Meisel, Z., 2014. Translating Research For Health Policy: Researchers' Perceptions And Use Of Social Media. *Health Affairs* 33, 1278–1285. <https://doi.org/10.1377/hlthaff.2014.0300>

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>

Grandjean, M., 2015. [Network analysis] Digital Humanities on Twitter, a small-world? URL <http://www.martingrandjean.ch/digital-humanities-on-twitter/> (accessed 1.19.16).

Grandjean, M., 2014. [DataViz] The digital humanities network on Twitter (#DH2014). Martin Grandjean. URL <http://www.martingrandjean.ch/dataviz-digital-humanities-twitter-dh2014/> (accessed 10.13.18).

Grandjean, M., 2013a. Digital Humanities Swiss Summer School on Twitter. Pegasus Data Project. URL <https://pegasusdata.com/2013/06/30/digital-humanities-swiss-summer-school-on-twitter/> (accessed 10.13.18).

Grandjean, M., 2013b. Vers une association francophone des humanités numériques ! Martin Grandjean. URL <http://www.martingrandjean.ch/association-francophone-humanites-numeriques/> (accessed 10.13.18).

Greenseid, L.O., 2008. Using Citation Analysis Methods to Assess the Influence of STEM Education Evaluation. ProQuest.

Gregory, K., Singh, S.S., 2018. Anger in Academic Twitter: Sharing, Caring, and Getting Mad Online. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 16, 176–193. <https://doi.org/10.31269/triplec.v16i1.890>

Griffith, B.C., 1989. Understanding Science: Studies of Communication and Information. *Communication Research* 16, 600–14.

Griffith, B.C., Mullins, N.C., 1972. Coherent Social Groups in Scientific Change. *Science*.

Groden, M., Kreiswirth, M., Szeman, I. (Eds.), 2005. *The Johns Hopkins guide to literary theory & criticism*, 2nd ed. ed. Johns Hopkins University Press, Baltimore, Md.

Gross, A.G., Harmon, J.E., Reidy, M.S., 2002a. *Communicating Science: The Scientific Article from the 17th Century to the Present*. Oxford University Press, USA.

Gross, A.G., Harmon, J.E., Reidy, M.S., 2002b. *Communicating science: the scientific article from the 17th century to the present*. Oxford Univ. Press, Oxford.

Grusin, R., 2013. The Dark Side of the Digital Humanities – Part 2 [WWW Document]. Thinking C21: Centre for 21st century studies. URL <http://www.c21uwm.com/2013/01/09/dark-side-of-the-digital-humanities-part-2/> (accessed 4.9.18).

Gruzd, A., Staves, K., Wilk, A., 2012. Connected scholars: Examining the role of social media in research practices of faculty using the UTAUT model. *Computers in Human Behavior* 28, 2340–2350. <https://doi.org/10.1016/j.chb.2012.07.004>

Guille, A., Favre, C., 2015. Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach. *Soc. Netw. Anal. Min.* 5, 18. <https://doi.org/10.1007/s13278-015-0258-0>

Guimera, R., Uzzi, B., Spiro, J., Amaral Luís A., N., 2005. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science* 308, 697–702. <https://doi.org/10.1126/science.1106340>

Hagmann, D., 2018. Reflections on the Use of Social Networking Sites as an Interactive Tool for Data Dissemination in Digital Archaeology. *Interdisciplinaria Archaeologica* 9.

Hagstrom, W.O., 1970. Factors Related to the Use of Different Modes of Publishing Research in Four Scientific Fields, in: *Communication among Scientists and Engineers*. Lexington Books, Lexington, pp. 87–124.

Haig, D., 2004. The Inexorable Rise of Gender and the Decline of Sex: Social Change in Academic Titles, 1945–2001. *Arch Sex Behav* 33, 87–96. <https://doi.org/10.1023/B:ASEB.0000014323.56281.0d>

Halliday, M.A.K., 1985. *An introduction to functional grammar*. E. Arnold, London ; Baltimore, Md., USA.

Hammond, R.A., Axelrod, R., 2006. The Evolution of Ethnocentrism. *Journal of Conflict Resolution* 50, 926–936. <https://doi.org/10.1177/0022002706293470>

Hara, N., Solomon, P., Kim, S.-L., Sonnenwald, D.H., 2003. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors

that impact collaboration. *J. Am. Soc. Inf. Sci.* 54, 952–965. <https://doi.org/10.1002/asi.10291>

Harrison, R., 2010. Unique Benefits of Conference Attendance as a Method of Professional Development for LIS Professionals. *The Serials Librarian* 59, 263–270. <https://doi.org/10.1080/0361526X.2010.489353>

Harvey, L., Green, D., 1993. Defining Quality. *Assessment & Evaluation in Higher Education* 18, 9–34. <https://doi.org/10.1080/0260293930180102>

Hashemi, M., 2017. The Infrastructure Behind Twitter: Scale [WWW Document]. URL https://blog.twitter.com/engineering/en_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale.html (accessed 5.29.20).

Hayes, B., 2017. Belles lettres Meets Big Data [WWW Document]. *American Scientist*. URL <https://www.americanscientist.org/article/belles-lettres-meets-big-data> (accessed 4.5.19).

Hayles, K., 1999. *How we became posthuman: virtual bodies in cybernetics, literature, and informatics*. University of Chicago Press, Chicago, Ill.

Hayles, N.K., 2012. How We Think: Transforming Power and Digital Technologies, in: Berry, D.M. (Ed.), *Understanding Digital Humanities*. Palgrave Macmillan UK, London, pp. 42–66. https://doi.org/10.1057/9780230371934_3

He, T., 2009. International scientific collaboration of China with the G7 countries. *Scientometrics* 80, 571–582. <https://doi.org/10.1007/s11192-007-2043-y>

Head, A.J., 2008. Information Literacy from the Trenches: How Do Humanities and Social Science Majors Conduct Academic Research? *CRL* 69, 427–446. <https://doi.org/10.5860/crl.69.5.427>

Hellqvist, B., 2009. Referencing in the humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology* 310–318. <https://doi.org/10.1002/asi.21256>

Henny-Krahmer, U., Sahle, P., 2018. Einreichungen zur DHd 2018 [WWW Document]. DHd-Blog. URL <https://dhd-blog.org/?p=9001> (accessed 10.13.18).

Henriksen, D., 2016. The rise in co-authorship in the social sciences (1980–2013). *Scientometrics* 107, 455–476. <https://doi.org/10.1007/s11192-016-1849-x>

Hernandez, G.I., 2016. The Responsibility of the International Legal Academic: Situating the Grammarians within the “Invisible College,” in: Nollkaemper, A., Werner, W., d'Aspremont, J., Gazzini, T. (Eds.), *International Law as a Profession*. Cambridge University Press, Cambridge.

Hicks, D., 2005. The Four Literatures of Social Science, in: Moed, H.F., Glänzel, W., Schmoch, U. (Eds.), *Handbook of Quantitative Science and Technology Research*. Kluwer Academic Publishers, Dordrecht, pp. 473–496. https://doi.org/10.1007/1-4020-2755-9_22

Hicks, D., Wang, J., 2009. Towards a Bibliometric Database for the Social Sciences and Humanities.

Hill, D.B., Willoughby, B.L.B., 2005. The Development and Validation of the Genderism and Transphobia Scale. *Sex Roles* 53, 531–544. <https://doi.org/10.1007/s11199-005-7140-x>

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* 102, 16569–16572. <https://doi.org/10.1073/pnas.0507655102>

Hjørland, B., 2000. Library and information science: practice, theory, and philosophical basis. *Information Processing & Management* 36, 501–531. [https://doi.org/10.1016/S0306-4573\(99\)00038-2](https://doi.org/10.1016/S0306-4573(99)00038-2)

Hockey, S., 2004. The History of Humanities Computing, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A Companion to Digital Humanities*. Blackwell Publishing Ltd, Malden, MA, USA, pp. 1–19.

Hogan, B., Quan-Haase, A., 2010. Persistence and Change in Social Media. *Bulletin of Science, Technology & Society* 30, 309–315.

Holmberg, K., Thelwall, M., 2014. Disciplinary Differences in Twitter Scholarly Communication. *Scientometrics* 101, 1027–1042. <https://doi.org/10.1007/s11192-014-1229-3>

Holmes, D.I., 1994. Authorship attribution. *Comput Hum* 28, 87–106. <https://doi.org/10.1007/BF01830689>

Holmes, D.I., Forsyth, R.S., 1995. The Federalist Revisited: New Directions in Authorship Attribution. *Literary and Linguistic Computing* 10, 111–127. <https://doi.org/10.1093/lrc/10.2.111>

Holton, R., 2000. Globalization's Cultural Consequences. *The ANNALS of the American Academy of Political and Social Science* 570, 140–152. <https://doi.org/10.1177/000271620057000111>

Honey, C., Herring, S.C., 2009. Beyond Microblogging: Conversation and Collaboration via Twitter, in: 2009 42nd Hawaii International Conference on System Sciences. Presented at the 2009 42nd Hawaii International Conference on System Sciences, IEEE, Waikoloa, Hawaii, USA, pp. 1–10. <https://doi.org/10.1109/HICSS.2009.89>

Hoover, D.L., 2004a. Testing Burrows's Delta. *Literary and Linguistic Computing* 19, 453–475. <https://doi.org/10.1093/lrc/19.4.453>

Hoover, D.L., 2004b. Delta Prime? *Literary and Linguistic Computing* 19, 477–495. <https://doi.org/10.1093/lrc/19.4.477>

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. <https://doi.org/10.1073/pnas.0307750100>

Howcroft, D., Trauth, E.M., 2005. *Handbook of Critical Information Systems Research: Theory and Application*. Edward Elgar Publishing.

Hsiao, T.-M., Chen, K., 2017. Yet another method for author co-citation analysis: A new approach based on paragraph similarity: Yet Another Method for Author Co-citation Analysis: A New Approach Based on Paragraph Similarity. *Proceedings of the Association for Information Science and Technology* 54, 170–178. <https://doi.org/10.1002/pra2.2017.14505401019>

Hudson, J., 1996. Trends in Multi-Authored Papers in Economics. *Journal of Economic Perspectives* 10, 153–158. <https://doi.org/10.1257/jep.10.3.153>

Huggett, J., 2012. Core or Periphery? Digital Humanities from an Archaeological Perspective. *Historical Social Research / Historische Sozialforschung* 37, 86–105.

Hughes, L., Constantopoulos, P., Dallas, C., 2015. Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A New Companion to Digital Humanities*. John Wiley & Sons, Ltd, Chichester, UK, pp. 150–170. <https://doi.org/10.1002/9781118680605.ch11>

Hyland, K., 1999. Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics* 20, 341–367. <https://doi.org/10.1093/applin/20.3.341>

Ibarra, H., 1992. Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm. *Administrative Science Quarterly* 37, 422–447.

ICWSM19, 2018. The 11th International AAAI Conference on Web and Social Media [WWW Document]. ICWSM-19 website. URL <https://www.icwsm.org/2019/> (accessed 11.16.18).

Ingram, S., Parker, A., 2002. Gender and Modes of Collaboration in an Engineering Classroom: A Profile of Two Women on Student Teams. *Journal of Business and Technical Communication* 16, 33–68. <https://doi.org/10.1177/1050651902016001002>

Ingvarsdóttir, H., Arnbjörnsdóttir, B., 2013. ELF and academic writing: a perspective from the Expanding Circle. *Journal of English as a Lingua Franca* 2. <https://doi.org/10.1515/jelf-2013-0006>

Inversini, A., Sage, R., Williams, N., Buhalis, D., 2015. The Social Impact of Events in Social Media Conversation, in: Tussyadiah, I., Inversini, A. (Eds.), *Information and Communication Technologies in Tourism 2015*. Springer International Publishing, Cham, pp. 283–294. https://doi.org/10.1007/978-3-319-14343-9_21

Ioannidis, J.P.A., 2015. A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *Journal of Psychosomatic Research* 78, 7–11. <https://doi.org/10.1016/j.jpsychores.2014.11.008>

Java, A., Song, X., Finin, T., Tseng, B., 2007. Why we twitter: understanding microblogging usage and communities. ACM Press, pp. 56–65. <https://doi.org/10.1145/1348549.1348556>

Jenders, M., Kasneci, G., Naumann, F., 2013. Analyzing and predicting viral tweets, in: *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*. Presented at the the 22nd International Conference, ACM Press, Rio de Janeiro, Brazil, pp. 657–664. <https://doi.org/10.1145/2487788.2488017>

Jensen, K.E., 2014. Linguistics in the digital humanities: (computational) corpus linguistics. *MedieKultur* 30, 20. <https://doi.org/10.7146/mediekultur.v30i57.15968>

Jeong, Y.K., Song, M., 2016. Applying content-based similarity measure to author co-citation analysis. <https://doi.org/10.9776/16212>

Jones, B., Dillman, K., Tang, R., Tang, A., Sharlin, E., Oehlberg, L., Neustaedter, C., Bateman, S., 2016. Elevating Communication, Collaboration, and Shared Experiences in Mobile Video through Drones. ACM Press, pp. 1123–1135. <https://doi.org/10.1145/2901790.2901847>

Jones, E., Lansey, J., Diedrich, F., 2012. Performance Measures for Improved Submarine Decision Making. APTIMA INC WOBURN MA.

Jones, S.E., 2016. Roberto Busa, S.J. and the emergence of humanities computing: the priest and the punched cards. Routledge, New York.

Jones, S.E., Nyhan, J., Rockwell, G., Sinclair, S., Terras, M., 2017. Reverse Engineering the First Humanities Computing Center. Presented at the Digital Humanities 2017, Montreal.

Kanwal, J., Smith, K., Culbertson, J., Kirby, S., 2017. Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition* 165, 45–52. <https://doi.org/10.1016/j.cognition.2017.05.001>

Kay, R., Lauricella, S., 2015. Investigating and Comparing Communication Media Used in Higher Education. *Journal of Communication Technology and Human Behaviors*. <https://doi.org/10.7726/jcthb.2015.1001>

Kemman, M., 2018. DHBenelux 2018 submissions. Max Kemman. URL <https://www.maxkemman.nl/2018/06/dhbenelux-2018-submissions/> (accessed 10.13.18).

Kemman, M., 2017. DHBenelux 2017 submissions. Max Kemman. URL <https://www.maxkemman.nl/2017/04/dhbenelux-2017-submissions/> (accessed 10.13.18).

Kemman, M., 2016a. DHBenelux submissions 2014-2016. Max Kemman. URL <https://www.maxkemman.nl/2016/04/dhbenelux-submissions-2014-2016/> (accessed 10.13.18).

Kemman, M., 2016b. DHBenelux 2016 submissions. Max Kemman. URL <https://www.maxkemman.nl/2016/02/dhbenelux-2016-submissions/> (accessed 10.13.18).

Kenny, S., 2015. About #transformDH. #TransformDH. URL <https://transformdh.org/about-transformdh/> (accessed 3.19.19).

Kim, H.J., Jeong, Y.K., Song, M., 2016. Content- and proximity-based author co-citation analysis using citation sentences. *Journal of Informetrics* 10, 954–966. <https://doi.org/10.1016/j.joi.2016.07.007>

Kim, T., Chang, A., Holland, L., Pentland, A. (Sandy), 2008. Meeting mediator: enhancing group collaboration with sociometric feedback, in: Proceeding of the Twenty-Sixth Annual CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI '08. Presented at the Proceeding of the twenty-sixth annual CHI conference extended abstracts, ACM Press, Florence, Italy, p. 3183. <https://doi.org/10.1145/1358628.1358828>

Kimmons, R., Veletsianos, G., 2016. Education scholars' evolving uses of twitter as a conference backchannel and social commentary platform: Twitter backchannel use. *Br J Educ Technol* 47, 445–464. <https://doi.org/10.1111/bjet.12428>

Kirsch, A., 2014. Technology is taking over english departments. The new republic.

Kirschenbaum, M., 2014. "What Is "Digital Humanities," and Why Are They Saying Such Terrible Things about It?. *differences* 25, 46–63.

Kirschenbaum, M., 2010. What is digital humanities and what's it doing in english departments? *ADE Bulletin* 150, 1–7.

Kirschenbaum, M.G., 2008. Mechanisms: new media and the forensic imagination. MIT Press, Cambridge, Mass.

Klavans, R., Boyack, K.W., 2006. Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology* 57, 251–263. <https://doi.org/10.1002/asi.20274>

Klein, J.T., 2015. *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field*. University of Michigan Press.

Koerner, K., 1989. Leonard Bloomfield and the "Cours de Linguistique Générale." *Cahiers Ferdinand de Saussure* 55–63.

Koh, A., 2012. The Challenges of Digital Scholarship [WWW Document]. *The Chronicle of Higher Education*. URL <https://www.chronicle.com/blogs/profhacker/the-challenges-of-digital-scholarship/38103> (accessed 4.9.19).

Koku, E., Nazer, N., Wellman, B., 2001. Netting Scholars: Online and Offline. *American Behavioral Scientist* 44, 1752–1774. <https://doi.org/10.1177/00027640121958023>

Koltay, T., 2016. Library and information science and the digital humanities: Perceived and real strengths and weaknesses. *Journal of Documentation* 72, 781–792. <https://doi.org/10.1108/JDOC-01-2016-0008>

Koschützki, D., Lehmann, K.A., Peeters, L., Richter, S., Tenfelde-Podehl, D., Zlotowski, O., 2005. Centrality Indices, in: Brandes, U., Erlebach, T. (Eds.), *Network Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 16–61. https://doi.org/10.1007/978-3-540-31955-9_3

Köseoglu, M.A., Okumus, F., Putra, E.D., Yildiz, M., Dogan, I.C., 2018. Authorship Trends, Collaboration Patterns, and Co-Authorship Networks in Lodging Studies (1990–2016). *Journal of Hospitality Marketing & Management* 27, 561–582. <https://doi.org/10.1080/19368623.2018.1399192>

Kozlowski, S.W.J., Bell, B.S., 2003. Work Groups and Teams in Organizations, in: Weiner, I.B. (Ed.), *Handbook of Psychology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. wei1214. <https://doi.org/10.1002/0471264385.wei1214>

Kraus, K.M., 2013. Alt-Research for Humanities PhDs [WWW Document]. Blog. URL <http://www.karikraus.com/?p=234> (accessed 8.8.19).

Kretschmer, H., Aguillo, I.F., 2005. New indicators for gender studies in Web networks. *Information Processing & Management* 41, 1481–1494. <https://doi.org/10.1016/j.ipm.2005.03.009>

Krieger, D.J., Belliger, A., 2014. *Interpreting networks: hermeneutics, actor-network theory & new media, Digital society*. Transcript, Bielefeld.

Krippendorff, K., 2018. Content analysis: an introduction to its methodology, Fourth Edition. ed. SAGE, Los Angeles.

Krishnan, A., 2009. What are Academic Disciplines? University of Southampton, NCRM E Prints Repository eprints.

Kronick, D.A., 2001. The Commerce of Letters: Networks and “Invisible Colleges” in Seventeenth- and Eighteenth-Century Europe. *The Library Quarterly* 71, 28–43. <https://doi.org/10.1086/603239>

Kuhn, P., Villeval, M.C., 2015. Are Women More Attracted to Co-operation Than Men? *Econ J* 125, 115–140. <https://doi.org/10.1111/ecoj.12122>

Kumar, S., 2018. Ethical Concerns in the Rise of Co-Authorship and Its Role as a Proxy of Research Collaborations. *Publications* 6, 37. <https://doi.org/10.3390/publications6030037>

Kumar, S., 2015. Co-authorship networks: a review of the literature. *Aslib Journal of Info Mgmt* 67, 55–73. <https://doi.org/10.1108/AJIM-09-2014-0116>

Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter, a social network or a news media? ACM Press, pp. 591–600. <https://doi.org/10.1145/1772690.1772751>

Kyvik, S., Teigen, M., 1996. Child Care, Research Collaboration, and Gender Differences in Scientific Productivity. *Science, Technology, & Human Values* 21, 54–71. <https://doi.org/10.1177/016224399602100103>

Lagemann, E.C., 1989. The Plural Worlds of Educational Research. *History of Education Quarterly* 29, 185. <https://doi.org/10.2307/368309>

Larivière, V., Gingras, Y., Archambault, É., 2006. Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics* 68, 519–533. <https://doi.org/10.1007/s11192-006-0127-8>

Larivière, V., Ni, C., Gingras, Y., Cronin, B., Sugimoto, C.R., 2013. Bibliometrics: Global gender disparities in science. *Nature News* 504, 211. <https://doi.org/10.1038/504211a>

Larivière, V., Vignola-Gagné, E., Villeneuve, C., Gélinas, P., Gingras, Y., 2011. Sex differences in research funding, productivity and impact: an analysis of Québec university professors. *Scientometrics* 87, 483–498. <https://doi.org/10.1007/s11192-011-0369-y>

Lasorsa, D., 2012. Transparency And Other Journalistic Norms On Twitter: The Role Of Gender. *Journalism Studies* 13, 402–417. <https://doi.org/10.1080/1461670X.2012.657909>

Lawrence, P.A., 2006. Men, Women, and Ghosts in Science. *PLoS Biology* 4, e19. <https://doi.org/10.1371/journal.pbio.0040019>

Layman, D., Elliott, T., 2019. Collaborative Authorship Trends in AMTA-Published Journals, 1988–2017. *Music Therapy Perspectives* 37, 93–99. <https://doi.org/10.1093/mtp/miy026>

Leahey, E., 2006. Gender Differences in Productivity: Research Specialization as a Missing Link. *Gender & Society* 20, 754–780. <https://doi.org/10.1177/0891243206293030>

Leane, E., Fletcher, L., Garg, S., 2019. Co-authorship trends in English literary studies, 1995–2015. *Studies in Higher Education* 44, 786–798. <https://doi.org/10.1080/03075079.2017.1405256>

Leaper, C., 1991. Influence and Involvement in Children's Discourse: Age, Gender, and Partner Effects. *Child Development* 62, 797. <https://doi.org/10.2307/1131178>

Lee, M.K., Yoon, H.Y., Smith, M., Park, H.J., Park, H.W., 2017. Mapping a Twitter scholarly communication network: a case of the association of internet researchers' conference. *Scientometrics* 112, 767–797. <https://doi.org/10.1007/s11192-017-2413-z>

Leech, G., 2014. The state of the art in corpus linguistics, in: Aijmer, K., Altenberg, B. (Eds.), *English Corpus Linguistics*. Taylor & Francis Group. <https://doi.org/10.4324/9781315845890-11>

Levenberg, L., Neilson, T., Rheams, D., 2018. Research methods for the digital humanities.

Lewin, K., 1936. Pyschology of Success and Failure. *Occupations: The Vocational Guidance Journal* 14, 926–930. <https://doi.org/10.1002/j.2164-5892.1936.tb00275.x>

Leydesdorff, L., Ophof, T., 2010. Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology* 61, 2365–2369. <https://doi.org/10.1002/asi.21371>

Leydesdorff, L., Salah, A.A.A., 2010. Maps on the Basis of the Arts & Humanities Citation Index: The Journals Leonardo and Art Journal Versus "Digital Humanities" as a Topic. *J. Am. Soc. Inf. Sci. Technol.* 61, 787–801. <https://doi.org/10.1002/asi.21303>

Li, E.Y., Liao, C.H., Yen, H.R., 2013. Co-authorship networks and research impact: A social capital perspective. *Research Policy* 42, 1515–1530. <https://doi.org/10.1016/j.respol.2013.06.012>

Library Carpentry, 2020. Library Carpentry [WWW Document]. Library Carpentry. URL <https://librarycarpentry.org/index.html> (accessed 4.4.20).

Lieberson, S., 1969. Measuring Population Diversity. *American Sociological Review* 34, 850. <https://doi.org/10.2307/2095977>

Lievrouw, L.A., 1990. Reconciling Structure and Process in the Study of Scholarly Communication, in: *Scholarly Communication and Bibliometrics*. pp. 59–69.

Lillis, T., Curry, M.J., 2006. Professional Academic Writing by Multilingual Scholars: Interactions With Literacy Brokers in the Production of English-Medium Texts. *Written Communication* 23, 3–35. <https://doi.org/10.1177/0741088305283754>

Lim, K.H., Datta, A., 2016. An interaction-based approach to detecting highly interactive Twitter communities using tweeting links. *WEB* 14, 1–15. <https://doi.org/10.3233/WEB-160328>

Lindsey, D., 1980. Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship. *Social Studies of Science* 10, 145–162.

Liu, A., 2016. Is Digital Humanities a Field? An Answer from the Point of View of Language. *Journal of Siberian Federal University. Humanities & Social Sciences* 9, 1546–1552. <https://doi.org/10.17516/1997-1370-2016-9-7-1546-1552>

Liu, A., 2013. The Meaning of the Digital Humanities. *PMLA-Publ. Mod. Lang. Assoc. Am.* 128, 409–423.

Liu, A., 2012a. The state of the digital humanities: A report and a critique. *Arts and Humanities in Higher Education* 11, 8–41. <https://doi.org/10.1177/1474022211427364>

Liu, A., 2012b. Where Is Cultural Criticism in the Digital Humanities?, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. pp. 490–509.

Liu, X., Bollen, J., Nelson, M.L., Van de Sompel, H., 2005. Co-authorship networks in the digital library research community. *Information Processing & Management* 41, 1462–1480. <https://doi.org/10.1016/j.ipm.2005.03.012>

Long, J.S., 1992. Measures of Sex Differences in Scientific Productivity. *Social Forces* 71, 159–178. <https://doi.org/10.1093/sf/71.1.159>

Losh, E., Wernimont, J., 2019. *Bodies of Information Intersectional Feminism and the Digital Humanities*. University of Minnesota Press, Minneapolis.

Lothian, A., Phillips, A., 2013. Can Digital Humanities Mean Transformative Critique? *Journal of e-Media Studies* 3. <https://doi.org/10.1349/PS1.1938-6060.A.425>

Lu, H., Feng, Y., 2009. A measure of authors' centrality in co-authorship networks based on the distribution of collaborative relationships. *Scientometrics* 81, 499–511. <https://doi.org/10.1007/s11192-008-2173-x>

Lubek, I., Innis, N.K., Kroger, R.O., McGuire, G.R., Stam, H.J., Herrmann, T., 1995. Faculty genealogies in five Canadian Universities: Historic-graphical and pedagogical concerns. *Journal of the History of the Behavioral Sciences* 31, 52–72. [https://doi.org/10.1002/1520-6696\(199501\)31:1<52::AID-JHBS2300310104>3.0.CO;2-B](https://doi.org/10.1002/1520-6696(199501)31:1<52::AID-JHBS2300310104>3.0.CO;2-B)

Lundberg, J., Tomson, G., Lundkvist, I., Sk?r, J., Brommels, M., 2006. Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics* 69, 575–589. <https://doi.org/10.1007/s11192-006-0170-5>

Machida, S., 2012. Does Globalization Render People More Ethnocentric? *Globalization and People's Views on Cultures: Does Globalization Render People More Ethnocentric?* *American Journal of Economics and Sociology* 71, 436–469. <https://doi.org/10.1111/j.1536-7150.2012.00835.x>

MacRoberts, M.H., MacRoberts, B.R., 1996. Problems of citation analysis. *Scientometrics* 36, 435–444.

Mahony, S., 2018. Cultural Diversity and the Digital Humanities. *Fudan Journal of the Humanities and Social Sciences*. <https://doi.org/10.1007/s40647-018-0216-0>

Mahony, S., Gao, J., 2018. Linguistic and Cultural Hegemony in the Digital Humanities, in: *Proceedings of the Digital Humanities Congress 2018*. Presented at the Sheffield: The Digital Humanities Institute, 2019, Sheffield, UK.

Mair, J., Thompson, K., 2009. The UK association conference attendance decision-making process. *Tourism Management* 30, 400–409. <https://doi.org/10.1016/j.tourman.2008.08.002>

Mäkelä, E., Tolonen, M., 2018. DHN2018 -An Analysis of a digital humanities conference, in: DHN 2018 : Digital Humanities in the Nordic Countries 3rd Conference: Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference, Helsinki, Finland, March 7-9, 2018. Presented at the Digital Humanities in the Nordic Countries, Rheinisch-Westfälische Technische Hochschule Aachen.

Makice, K., 2009. Twitter API: up and running, 1st ed. ed. O'Reilly, Sebastopol, Calif.

Mali, F., Kronegger, L., Doreian, P., Ferligoj, A., 2012. Dynamic Scientific Co-Authorship Networks, in: Scharnhorst, A., Börner, K., van den Besselaar, P. (Eds.), *Models of Science Dynamics*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 195–232. https://doi.org/10.1007/978-3-642-23068-4_6

Mannion, C.M., 2018. Research Guides: Global China Studies: Digital Chinese Humanities [WWW Document]. URL <https://guides.nyu.edu/dh> (accessed 11.8.19).

Manovich, L., 2001. *The language of new media*, Leonardo. MIT Press, Cambridge, Mass.

Marsden, P.V., 2002. Egocentric and sociocentric measures of network centrality. *Social Networks* 24, 407–422. [https://doi.org/10.1016/S0378-8733\(02\)00016-3](https://doi.org/10.1016/S0378-8733(02)00016-3)

Massanari, A., 2017. #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society* 19, 329–346. <https://doi.org/10.1177/1461444815608807>

Matarese, V., 2013. Supporting research writing in non-anglophone Europe: reflections and recurring themes, in: *Supporting Research Writing*. Elsevier, pp. 257–269. <https://doi.org/10.1016/B978-1-84334-666-1.50025-4>

Mattsson, P., Sundberg, C.J., Laget, P., 2011. Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics* 87, 99–105. <https://doi.org/10.1007/s11192-010-0310-9>

Matzat, U., 2004. Academic communication and Internet Discussion Groups: transfer of information or creation of social contacts? *Social Networks* 26, 221–255. <https://doi.org/10.1016/j.socnet.2004.04.001>

McCain, K.W., 1990. Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science* 41, 433–443. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<433::AID-ASI11>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<433::AID-ASI11>3.0.CO;2-Q)

McCarty, W., 2015. Becoming Interdisciplinary, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A New Companion to Digital Humanities*. John Wiley & Sons, Ltd, Chichester, UK, pp. 67–83. <https://doi.org/10.1002/9781118680605.ch5>

McCarty, W., 2011. Beyond chronology and profession: Discovering how to write a history of the digital humanities, in: Paper given at Hidden Histories:

Symposium on Methodologies for the History of Computing in the Humanities, c.1949–1980. University College London.

McCarty, W., 2008. What's going on? *Literary and Linguistic Computing* 23, 253–61. <https://doi.org/10.1093/lcl/fqn014>

McCarty, W., 2006. Tree, turf, centre, archipelago—or wild acre? Metaphors and stories for humanities computing. *Literary and Linguistic Computing* 21, 1–13.

McCarty, W., 2003a. *Humanities Computing*. Palgrave Macmillan, Basingstoke.

McCarty, W., 2003b. Humanities Computing, in: *Encyclopedia of Library and Information Science*. Marcel Dekker, New York.

McCarty, W., 1999. Humanities computing as interdiscipline [WWW Document]. URL <http://www.iath.virginia.edu/hcs/mccarty.html> (accessed 4.9.18).

McCarty, W., 1992. HUMANIST: Lessons from a Global Electronic Seminar. Computers and the Humanities 26, 205–222. <https://doi.org/10.1007/BF00058618>

McClure, D., 2014. Visualizing 27 years, 12 million words of the Humanist list. *david mcClure*. URL <http://dclure.org/essays/visualizing-the-humanist/> (accessed 10.13.18).

McFadyen, M.A., Cannella, A.A., 2004. Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange Relationships. *Academy of Management Journal* 47, 735–746. <https://doi.org/10.5465/20159615>

McGann, J., 2010. Sustainability: The Elephant in the Room [WWW Document]. Online Humanities Scholarship: The Shape of Things to Come. URL <https://cnx.org/contents/PVdH0-ID@1.3:AeBKUQq-@2/Sustainability-The-Elephant-in-the-Room> (accessed 12.1.18).

McGrath, C., Blythe, J., Krackhardt, D., 1997. The effect of spatial arrangement on judgments and errors in interpreting graphs. *Social Networks* 19, 223–242. [https://doi.org/10.1016/S0378-8733\(96\)00299-7](https://doi.org/10.1016/S0378-8733(96)00299-7)

McGregor, S.C., Mourão, R.R., 2016. Talking Politics on Twitter: Gender, Elections, and Social Networks. *Social Media + Society* 2, 205630511666421. <https://doi.org/10.1177/2056305116664218>

McMurtry, J., 1985. English language, English literature: the creation of an academic discipline. Mansell, London.

McPherson, T., 2012. Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. University of Minnesota Press, pp. 139–160. <https://doi.org/10.5749/minnesota/9780816677948.003.0017>

McPherson, T., 2009. Introduction: Media Studies and the Digital Humanities. *Cinema Journal* 48, 119–123.

Meadows, A.J. (Ed.), 1980. *Development of Science Publishing in Europe*. Elsevier Science Publishers ; sole distributors for the U.S.A. and Canada, Elsevier North-Holland, Amsterdam ; New York : New York, NY.

Meadows, A.J., 1974. *Communication in science*. Butterworths, London.

Meeks, E., 2011. Comprehending the Digital Humanities. *Digital Humanities Specialist*. URL <https://dhs.stanford.edu/comprehending-the-digital-humanities/> (accessed 10.29.18).

Mehra, A., Kilduff, M., Brass, D.J., 2001. The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance. *Administrative Science Quarterly* 46, 121. <https://doi.org/10.2307/2667127>

Melin, G., 1999. Impact of national size on research collaboration: A comparison between Northern European and American universities. *Scientometrics* 46, 161–170. <https://doi.org/10.1007/BF02766300>

Melin, G., Persson, O., 1996. Studying research collaboration using co-authorships. *Scientometrics* 36, 363–377. <https://doi.org/10.1007/BF02129600>

Menzel, H., 1968. Informal Communication in Science: Its Advantages and its Formal Analogues, in: Montgomery, E.B. (Ed.), *The Foundations of Access to Knowledge*. [Syracuse, N.Y.]Division of Summer Sessions, Syracuse University, pp. 153–163.

Merton, R.K., 1968. The Matthew Effect in Science: The reward and communication systems of science are considered. *Science* 159, 56–63. <https://doi.org/10.1126/science.159.3810.56>

Milkman, K.L., Chugh, D., Bazerman, M.H., 2009. How Can Decision Making Be Improved? *Perspect Psychol Sci* 4, 379–383. <https://doi.org/10.1111/j.1745-6924.2009.01142.x>

Miquel, J.F., Okubo, Y., 1994. Structure of international collaboration in science-part II: Comparisons of profiles in countries using a link indicator. *Scientometrics* 29, 271–297. <https://doi.org/10.1007/BF02017977>

Mitkov, R. (Ed.), 2014. *The Oxford Handbook of Computational Linguistics* 2nd edition, 2nd ed. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199573691.001.0001>

Moed, H.F., 2010. Measuring contextual citation impact of scientific journals. *Journal of Informetrics* 4, 265–277. <https://doi.org/10.1016/j.joi.2010.01.002>

Moed, H.F., 2006. *Citation Analysis in Research Evaluation*. Springer Science & Business Media.

Molina, B., 2019. Twitter overcounted active users since 2014, shares surge on profit hopes [WWW Document]. USA Today. URL <https://www.usatoday.com/story/tech/news/2017/10/26/twitter-overcounted-active-users-since-2014-shares-surge/801968001/> (accessed 10.25.18).

Moody, J., 2004. The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review* 69, 213–238. <https://doi.org/10.1177/000312240406900204>

Moravec, M., 2018. Exceptionalism in Digital Humanities: Community, Collaboration, and Consensus, in: Kim, D., Stommel, J. (Eds.), *Disrupting the Digital Humanities*. Punctum Books, Santa Barbara, CA, pp. 169–196.

Moreno, A., Redondo, T., 2016. Text Analytics: the convergence of Big Data and Artificial Intelligence. *IJIMAI* 3, 57–64.

Moreno, A.I., Rey-Rocha, J., Burgess, S., López-Navarro, I., Sachdev, I., 2012. Spanish researchers' perceived difficulty writing research articles for English-medium journals: the impact of proficiency in English versus publication experience [WWW Document]. Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos. URL <http://eprints.rclis.org/29319/> (accessed 1.23.19).

Moretti, F., 2005. Graphs, maps, trees: abstract models for literary history, Paperback edition. ed. Verso, London New York.

Myers, S.A., Sharma, A., Gupta, P., Lin, J., 2014. Information network or social network?: the structure of the twitter follow graph, in: Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion. Presented at the the 23rd International Conference, ACM Press, Seoul, Korea, pp. 493–498. <https://doi.org/10.1145/2567948.2576939>

Nadal, K.L., 2017. The SAGE encyclopedia of psychology and gender.

Nagasaki, K., Muller, A.C., 2012. Trends in Digital Scholarship in the Humanities in Japan, in: Digital Humanities Australasia Conference Abstracts. Presented at the Digital Humanities Australasia Conference, Australian National University.

National Science Foundation, 2017. 2016 Doctorate Recipients from U.S. Universities.

Neal, D.R., 2012. Social media for academics: a practical guide. Chandos Publishing, Oxford.

NEH, 2020. NEH Timeline [WWW Document]. National Endowment for the Humanities (NEH). URL <https://www.neh.gov/about/history/timeline> (accessed 4.27.20).

Nerbonne, J., Tonelli, S., 2016. Introduction to the Special Issue on Digital Humanities and Computational Linguistics. Italian Journal Of Computational Linguistics 2, 7–10.

Newman, M.E.J., 2005. A measure of betweenness centrality based on random walks. Social Networks 27, 39–54. <https://doi.org/10.1016/j.socnet.2004.11.009>

Newman, M.E.J., 2001. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences 98, 404–409. <https://doi.org/10.1073/pnas.98.2.404>

Nickerson, R.S., 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology 2, 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>

Nielsen Mobile, 2009. Nielsen Mobile Highlights Twitter and Predicto as Industry Leaders; Report Details the Diversity Among Free Versus Paid Mobile Communities [WWW Document]. PRWeb. URL <https://www.prweb.com/releases/2009/04/prweb2367764.htm> (accessed 10.26.18).

Noordenbos, G., 1992. Explanations for Differences in Publication Rates Between Male and Female Academics and Between Productive and Less Productive Women. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique 35, 22–45. <https://doi.org/10.1177/075910639203500102>

Northeastern University, 1986. The Women Writers Project [WWW Document]. The Women Writers Project. URL <https://www.wwp.northeastern.edu/> (accessed 1.23.19).

Nowviskie, B., 2016a. On the Origin of “Hack” and “Yack,” in: Debates in the Digital Humanities: 2016. University of Minnesota Press, Minneapolis London.

Nowviskie, B., 2016b. capacity through care [WWW Document]. Bethany Nowviskie. URL <http://nowviskie.org/2016/capacity-through-care/> (accessed 1.23.19).

Nowviskie, B., 2015. on capacity and care [WWW Document]. Bethany Nowviskie. URL <https://nowviskie.org/2015/on-capacity-and-care/> (accessed 1.23.19).

Noyons, E.C.M., Moed, H.F., Luwel, M., 1999. Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science* 50, 115–131. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:2<115::AID-ASI3>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-4571(1999)50:2<115::AID-ASI3>3.0.CO;2-J)

Nyhan, J., 2016. In search of identities in the digital humanities: The early history of Humanist, in: Social Media Archaeology and Poetics. MIT Press, Cambridge, MA.

Nyhan, J., Duke-Williams, O., 2014a. Joint and multi-authored publication patterns in the Digital Humanities. *Literary and Linguistic Computing* 29, 387–399. <https://doi.org/10.1093/lrc/fqu018>

Nyhan, J., Duke-Williams, O., 2014b. Is Digital Humanities A Collaborative Discipline? Joint-authorship Patterns Clash with Defining Narrative. <http://blogs.lse.ac.uk/impactofsocialsciences/2014...<http://blogs.lse.ac.uk/impactofsocialsciences/2014/09/10/joint-authorship-digital-humanities-collaboration/>>. URL <http://blogs.lse.ac.uk/impactofsocialsciences/2014/09/10/joint-authorship-digital-humanities-collaboration/> (accessed 1.8.16).

Nyhan, J., Flinn, A., 2016. Computation and the Humanities: Towards an Oral History of Digital Humanities. Springer Berlin Heidelberg, New York, NY.

Nyhan, J., Passarotti, M., 2019. One Origin of Digital Humanities: Fr Roberto Busa in His Own Words.

Nyhan, J., Terras, M., 2017. Uncovering ‘hidden’ contributions to the history of Digital Humanities: the Index Thomisticus’ female keypunch operators. Presented at the Digital Humanities 2017, Montreal, Canada.

Nyhan, J., Terras, M., Vanhoutte, E., 2013. Introduction, in: Defining Digital Humanities: A Reader. Ashgate Publishing Ltd.

Nyhan, Julianne., Flinn, Andrew., Welsh, Anne., 2015. Oral History and the Hidden Histories project: towards histories of computing in the humanities. *Digital Scholarship in the Humanities* 30, 71–85. <https://doi.org/10.1093/lrc/fqt044>

Nylenna, M., Riis, P., Karlsson, Y., 1994. Multiple Blinded Reviews of the Same Two Manuscripts: Effects of Referee Characteristics and Publication Language. *JAMA* 272, 149. <https://doi.org/10.1001/jama.1994.03520020075021>

Nyström, J., 2005. The definition of partnering as a Wittgenstein family-resemblance concept. *Construction Management and Economics* 23, 473–481. <https://doi.org/10.1080/01446190500040026>

O'Donnell, D.P., Walter, K.L., Gil, A., Fraistat, N., 2015. Only Connect: The Globalization of the Digital Humanities, in: A New Companion to Digital Humanities. John Wiley & Sons, pp. 493–510.

Ogan, C., Varol, O., 2017. What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement: Twitter use during Gezi Park. *Information, Communication & Society* 20, 1220–1238. <https://doi.org/10.1080/1369118X.2016.1229006>

Ohio Humanities, 2020. About - Ohio Humanities [WWW Document]. Ohio Humanities. URL <http://www.ohiohumanities.org/about/> (accessed 4.27.20).

Okubo, Y., Miquel, J.F., Frigoletto, L., Doré, J.C., 1992. Structure of international collaboration in science: Typology of countries through multivariate techniques using a link indicator. *Scientometrics* 25, 321–351. <https://doi.org/10.1007/BF02028090>

Olson, C.M., 2002. Publication Bias in Editorial Decision Making. *JAMA* 287, 2825. <https://doi.org/10.1001/jama.287.21.2825>

Oppermann, M., Chon, K.-S., 1997. Convention participation decision-making process. *Annals of Tourism Research* 24, 178–191. [https://doi.org/10.1016/S0160-7383\(96\)00049-7](https://doi.org/10.1016/S0160-7383(96)00049-7)

Orduña-Malea, E., Torres-Salinas, D., Delgado López-Cózar, E., 2015. Hyperlinks embedded in twitter as a proxy for total external in-links to international university websites: Hyperlinks Embedded in Twitter as a Proxy for Total External In-Links to International University Websites. *J Assn Inf Sci Tec* 66, 1447–1462. <https://doi.org/10.1002/asi.23291>

Paasi, A., 2005. Globalisation, Academic Capitalism, and the Uneven Geographies of International Journal Publishing Spaces. *Environment and Planning A* 37, 769–789. <https://doi.org/10.1068/a3769>

Pal, K., Avery, N., Boston, P., Campagnolo, A., De Stefani, C., Matheson-Pollock, H., Panozzo, D., Payne, M., Schüller, C., Sanderson, C., Scott, C., Smith, P., Smither, R., Sorkine-Hornung, O., Stewart, A., Stewart, E., Stewart, P., Terras, M., Walsh, B., Ward, L., Yamada, L., Weyrich, T., 2016. Digitally reconstructing the Great Parchment Book: 3D recovery of fire-damaged historical documents. *Digital Scholarship Humanities* fqw057. <https://doi.org/10.1093/llc/fqw057>

Pan, R.K., Kaski, K., Fortunato, S., 2012. World citation and collaboration networks: uncovering the role of geography in science. *Scientific Reports* 2, 902.

Pannapacker, W., 2013. Stop Calling It “Digital Humanities”: And 9 Other Strategies to Help Liberal-arts Colleges Join the Movement. *The Chronicle of Higher Education*.

Parry, D., 2011. Using Twitter - But Not in the Classroom [WWW Document]. Learning Through Digital Media. URL <http://mcpress.media-commons.org/artoflearning/using-twitter%E2%80%94but-not-in-the-classroom/> (accessed 4.9.19).

Patience, G.S., Patience, C.A., Blais, B., Bertrand, F., 2017. Citation analysis of scientific categories. *Helion* 3, e00300. <https://doi.org/10.1016/j.heliyon.2017.e00300>

Peet, R., 2005. From Eurocentrism to Americentrism. *Antipode* 37, 936–943. <https://doi.org/10.1111/j.0066-4812.2005.00542.x>

Peñas, C.S., Willett, P., 2006. Brief communication: Gender differences in publication and citation counts in librarianship and information science research. *Journal of Information Science* 32, 480–485. <https://doi.org/10.1177/0165551506066058>

Peoples, B.K., Midway, S.R., Sackett, D., Lynch, A., Cooney, P.B., 2016. Twitter Predicts Citation Rates of Ecological Research. *PLOS ONE* 11, e0166570. <https://doi.org/10.1371/journal.pone.0166570>

Perianes-Rodriguez, A., Waltman, L., van Eck, N.J., 2016. Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics* 10, 1178–1195. <https://doi.org/10.1016/j.joi.2016.10.006>

Perkmann, M., Walsh, K., 2009. The two faces of collaboration: impacts of university-industry relations on public research. *Industrial and Corporate Change* 18, 1033–1065. <https://doi.org/10.1093/icc/dtp015>

Perry-Smith, J.E., Shalley, C.E., 2003. The Social Side of Creativity: A Static and Dynamic Social Network Perspective. *Academy of Management Review* 28, 89–106. <https://doi.org/10.5465/amr.2003.8925236>

Peters, H.P.F., van Raan, A.F.J., 1993. Co-word-based science maps of chemical engineering. Part II: Representations by combined clustering and multidimensional scaling. *Research Policy* 22, 47–71. [https://doi.org/10.1016/0048-7333\(93\)90032-D](https://doi.org/10.1016/0048-7333(93)90032-D)

Phillipson, R., 2012. Linguistic Imperialism, in: Chapelle, C.A. (Ed.), *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd, Oxford, UK, p. wbeal0718. <https://doi.org/10.1002/9781405198431.wbeal0718>

Pinsken, D., 2009. On Mendenhall and compelling evidence of Marlowe authorship by Daryl Pinsken. The Marlowe-Shakespeare Connection Blog. URL <http://marlowe-shakespeare.blogspot.ca/2009/02/onmendenhall-and-compelling-evidence.html> (accessed 10.5.18).

Pitman, T., Taylor, C., 2017. Where's the ML in DH? And Where's the DH in ML? The Relationship between Modern Languages and Digital Humanities, and an Argument for a Critical DHML. *DHQ* 011.

PMLA Editorial Board, 2018. “Varieties of Digital Humanities” (session description). Presented at the Program for MLA 2018.

Ponds, R., 2009. The limits to internationalization of scientific research collaboration. *J Technol Transf* 34, 76–94. <https://doi.org/10.1007/s10961-008-9083-1>

Ponomariov, B., Boardman, C., 2016. What is co-authorship? *Scientometrics* 109, 1939–1963. <https://doi.org/10.1007/s11192-016-2127-7>

Poole, A.H., 2017. The conceptual ecology of digital humanities. *Journal of Documentation* 73, 91–122. <https://doi.org/10.1108/JD-05-2016-0065>

Poole, A.H., Garwood, D.A., 2018. Interdisciplinary scholarly collaboration in data-intensive, public-funded, international digital humanities project work. *Library & Information Science Research* 40, 184–193. <https://doi.org/10.1016/j.lisr.2018.08.003>

Porter, A.L., Rafols, I., 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* 81, 719–745. <https://doi.org/10.1007/s11192-008-2197-2>

Posner, M., 2015. Selfies & Feminism | Selfies, Snapchat, & Cyberbullies. URL <http://miriamposner.com/dh150w15/selfies-feminism/> (accessed 1.23.19).

Pressman, J., Swanstrom, L., 2013. The Literary And/As the Digital Humanities. *DHQ: Digital Humanities Quarterly* 7.

Pulizzi, J., 2014. In the Near Future, Only Very Wealthy Colleges Will Have English Departments. *The New Republic*.

Puschmann, C., Bastos, M., 2015. How Digital Are the Digital Humanities? An Analysis of Two Scholarly Blogging Platforms. *PLoS One* 10, e0115035. <https://doi.org/10.1371/journal.pone.0115035>

Puschmann, C., Weller, K., E, D., 2011. Studying Twitter conversations as (dynamic) graphs: Visualization and structural comparison, in: *Proceedings of General Online Research*.

Quan-Haase, A., Martin, K., McCay-Peet, L., 2015a. Networks of Digital Humanities Scholars: The Informational and Social Uses and Gratifications of Twitter. *Big Data & Society* 2. <https://doi.org/10.1177/2053951715589417>

Quan-Haase, A., Martin, K., McCay-Peet, L., 2015b. Networks of Digital Humanities Scholars: The Informational and Social Uses and Gratifications of Twitter. *Big Data & Society* 2. <https://doi.org/10.1177/2053951715589417>

Raben, J., 1991. Humanities computing 25 years later. *Computers and the Humanities* 25, 341–350.

Raben, J., 1966. Prospect. *Computers and the Humanities* 1, 1–2.

Radicchi, F., Castellano, C., 2012. Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics* 6, 121–130. <https://doi.org/10.1016/j.joi.2011.09.002>

Rafols, I., Meyer, M., 2010. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics* 82, 263–287. <https://doi.org/10.1007/s11192-009-0041-y>

Rambukanna, N., 2015. FCJ-194 From #RaceFail to #Ferguson: The Digital Intimacies of Race-Activist Hashtag Publics. *The Fibreculture Journal* 160–189. <https://doi.org/10.15307/fcj.26.194.2015>

Ramsay, S., 2013a. On building, in: *Defining Digital Humanities: A Reader*. Ashgate Publishing, Surrey/Burlington, pp. 243–246.

Ramsay, S., 2013b. Who's in and who's out, in: *Defining Digital Humanities: A Reader*. Ashgate Publishing, Surrey/Burlington, pp. 239–242.

Ramsay, S., Rockwell, G., 2012. *Developing Things: Notes toward an Epistemology of Building in the Digital Humanities*. Univ Minnesota Press, Minneapolis.

Rangel, F., Rosso, P., 2013. Use of language and author profiling: Identification of gender and age. *Natural Language Processing and Cognitive Science* 2013, 177–186.

Rangel, F., Rosso, P., Montes-y-Gómez, M., Potthast, M., Stein, B., 2018. Overview of the 6th author profiling task at pan 2018: multimodal gender identification in Twitter. *Working Notes Papers of the CLEF (2018)*.

Ravallion, M., Wagstaff, A., 2011. On Measuring Scholarly Influence by Citations. *Scientometrics* 88, 321–337. <https://doi.org/10.1007/s11192-011-0375-0>

Reichelt, L., 2007. Ambient Intimacy. *Disambiguity* blog. URL <http://www.disambiguity.com/ambient-intimacy/> (accessed 10.25.18).

Reinbothe, R., 2013. Der Boykott gegen die deutschen Wissenschaftler und die deutsche Sprache nach dem Ersten Weltkrieg [The boycott against German scientists and the German language after World War I]. *Dtsch med Wochenschr* 138, 2685–2690. <https://doi.org/10.1055/s-0033-1359937>

Rhoten, D., Pfirman, S., 2007. Women in interdisciplinary science: Exploring preferences and consequences. *Research Policy* 36, 56–75. <https://doi.org/10.1016/j.respol.2006.08.001>

Rice, J., 2013. Occupying the Digital Humanities. *College English* 75, 360–378.

Risam, R., 2018. New digital worlds: postcolonial digital humanities in theory, praxis, and pedagogy. Northwestern University Press, Evanston, Illinois.

Risam, R., 2015a. Introduction: Gender, Globalization and the Digital. <https://doi.org/10.7264/n3rn364r>

Risam, R., 2015b. Toxic femininity 4.0. First Monday 20. <https://doi.org/10.5210/fm.v20i4.5896>

Robertson, S., 2016. The Differences between Digital History and Digital Humanities, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities: 2016*. U of Minnesota Press, pp. 289–307.

Robinson, L., Priego, E., Bawden, D., 2015. Library and information science and digital humanities: two disciplines, joint future?, in: Pehar, F., Schlogi, C., Wolff, C. (Eds.), *Re-Inventing Information Science in the Networked Society*. Presented at the 14th International Symposium on Information Science, Zadar, Croatia.

Rochat, Y., 2014. The DH 2014 Conference in Lausanne - A feedback. Yannick Rochat. URL <https://yro.ch/dh-2014-conference-lausanne-summary/> (accessed 10.28.18).

Rockenbach, B., 2013. Introduction. *Journal of Library Administration* 53, 1–9. <https://doi.org/10.1080/01930826.2013.756676>

Rockwell, Geoffrey, Nyhan, J., Welsh, A., Salmon, J., 2012a. Trading stories: An oral history conversation between Geoffrey Rockwell and Julianne Nyhan. *Digital Humanities Quarterly* 6.

Rockwell, Geoffrey, Organisciak, P., Meredith-Lobay, M., Ranaweera, K., Ruecker, S., Nyhan, J., 2012b. The Design of an International Social Media Event: A Day in the Life of the Digital Humanities. *DHQ* 006.

Rockwell, G., Organisciak, P., Meredith-Lobay, M., Ranaweera, K., Ruecker, S., Nyhan, J., 2012. The Design of an International Social Media Event: A Day in the Life of the Digital Humanities. DHQ 006.

Rockwell, G., Sinclair, S., 2016. The Swallow Flies Swiftly Through: An Analysis of Humanist, in: Hermeneutica: The Rhetoric of Text Analysis. MIT Press.

Rockwell, G., Sinclair, S., 2012. The Swallow Flies Swiftly Through: An Analysis of Humanist, in: DH2012. Presented at the Digital Humanities 2012, Hamburg, Germany.

Rockwell, G., Victoria, S., Sophia, H., Sean, G., Harvey, Q., 2011. Computing in Canada: A history of the incunabular years, in: Digital Humanities. Stanford University Library, Stanford, pp. 207–210.

Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J., 2014. Core-Periphery Structure in Networks. SIAM J. Appl. Math. 74, 167–190. <https://doi.org/10.1137/120881683>

Rørstad, K., Aksnes, D.W., 2015. Publication rate expressed by age, gender and academic position – A large-scale analysis of Norwegian academic staff. Journal of Informetrics 9, 317–333. <https://doi.org/10.1016/j.joi.2015.02.003>

Roser, M., Ritchie, H., Ortiz-Ospina, E., 2020. World Population Growth. Our World in Data.

Ross, C., 2012. Social media for digital humanities and community engagement, in: Digital Humanities in Practice. United Kingdom: Facet Publishing, pp. 23–46.

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>

Rowlands, I., Nicholas, D., Russell, B., Canty, N., Watkinson, A., 2011. Social media use in the research workflow. Learned Publishing 24, 183–195. <https://doi.org/10.1087/20110306>

Salah, A.A.A., Leydesdorff, L., Scharnhorst, A., 2010. Mapping the Flow of “Digital Humanities,” in: DH2010. Presented at the Digital Humanities 2010, London, UK.

Salah, A.A.A., Scharnhorst, A., Wyatt, S., 2015. Analysing an Academic Field through the Lenses of Internet Science: Digital Humanities as a Virtual Community, in: Tiropanis, T., Vakali, A., Sartori, L., Burnap, P. (Eds.), Internet Science, Lecture Notes in Computer Science. Springer International Publishing, pp. 78–89.

Salas, E., Cooke, N.J., Rosen, M.A., 2008. On Teams, Teamwork, and Team Performance: Discoveries and Developments. Hum Factors 50, 540–547. <https://doi.org/10.1518/001872008X288457>

Sample, M., 2011. The Digital Humanities is not about Building, it's about Sharing. Sample Reality. URL <http://www.samplereality.com/2011/05/25/the-digital-humanities-is-not-about-building-its-about-sharing/> (accessed 4.17.18).

Sandelin, B., Sarafoglou, N., 2004. Language and scientific publication statistics. LPLP 28, 1–10. <https://doi.org/10.1075/lplp.28.1.02san>

Sandstrom, P.E., 1998. Information Foraging Among Anthropologists in the Invisible College of Human Behavioral Ecology: An Author Co-citation Analysis. Indiana University.

Schmoch, U., Schubert, T., 2008. Are international co-publications an indicator for quality of scientific research? *Scientometrics* 74, 361–377. <https://doi.org/10.1007/s11192-007-1818-5>

Schnapp, J., Presner, T., 2009. The Digital Humanities Manifesto 2.0. UCLA Mellon Seminar in Digital Humanities. URL http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf (accessed 1.10.16).

Schreibman, S., Siemens, R., Unsworth, J. (Eds.), 2004a. *A Companion to Digital Humanities*. Blackwell Publishing Ltd, Oxford, UK.

Schreibman, S., Siemens, R., Unsworth, J., 2004b. The Digital Humanities and Humanities Computing: An Introduction, in: *Companion to Digital Humanities*, Blackwell Companions to Literature and Culture. Blackwell Publishing Professional, Oxford.

Schreibman, S., Siemens, R.G., Unsworth, J. (Eds.), 2016. *A New Companion to Digital Humanities*. John Wiley & Sons Inc, Chichester, West Sussex, UK.

Schroeder-Gudehus, B., 1990. Internationale Wissenschaftsbeziehungen und auswärtige Kulturpolitik 1919-1933. Vom Boykott und Gegen-Boykott zu ihrer Wiederaufnahme. Vierhaus, Rudolf/Brocke, Bernhard vom (Hg.): *Forschung im Spannungsfeld von Politik und Gesellschaft. Geschichte und Struktur der Kaiser-Wilhelm-/Max-Planck-Gesellschaft*. Stuttgart. S 858–885.

Seglen, P.O., 1997. Why the impact factor of journals should not be used for evaluating research. *BMJ* 314, 497–497. <https://doi.org/10.1136/bmj.314.7079.497>

Sehl, K., 2020. How the Twitter Algorithm Works in 2020 and How to Make it Work for You [WWW Document]. Hootsuite Social Media Management. URL <https://blog.hootsuite.com/twitter-algorithm/> (accessed 5.29.20).

Severt, D., Wang, Y., Chen, P.-J., Breiter, D., 2007. Examining the motivation, perceived performance, and behavioral intentions of convention attendees: Evidence from a regional conference. *Tourism Management* 28, 399–408. <https://doi.org/10.1016/j.tourman.2006.04.003>

Sharing is Caring, 2019. About. Sharing is Caring. URL <http://sharecare.nu/about/> (accessed 4.4.20).

Shewan, L.G., Coats, A.J.S., Glatz, J.A., Bennett, C.C., 2005. Contemporary (post-Wills) survey of the views of Australian medical researchers: importance of funding, infrastructure and motivators for a research career. *Medical Journal of Australia* 183, 606–611. <https://doi.org/10.5694/j.1326-5377.2005.tb00051.x>

Siemens, L., 2009. “It’s a team if you use ‘reply all’”: An exploration of research teams in digital humanities environments. *Lit. Linguist. Comput.* 24, 225–233. <https://doi.org/10.1093/linc/fqp009>

Siemens, R., 2016. Preface: Communities of practice, the methodological commons, and digital self-determination in the humanities, in: Crompton, C., Lane, R.J.,

Siemens, R.G. (Eds.), *Doing Digital Humanities: Practice, Training, Research*. Routledge, New York, NY, pp. xxi–xxxiii.

Silverman, R., 2013. Mary Beard: Twitter bomb threat was “scary.”

Skene, E., 2019. Research Guides: Digital Humanities: Digital Humanities Tools and Methods [WWW Document]. URL <https://researchguides.wcu.edu/digitalhumanities/toolsmethods> (accessed 8.8.19).

Smart, J.C., Bayer, A.E., 1986. Author collaboration and impact: A note on citation rates of single and multiple authored articles. *Scientometrics* 10, 297–305. <https://doi.org/10.1007/BF02016776>

Snyder, E.E., 2014. A Framework for Supporting the Digital Humanities: An Alternative to the DH Centre, in: Mills, C., Pidd, M., Ward, E. (Eds.), *Proceedings of the Digital Humanities Congress 2012*. Presented at the Digital Humanities Congress 2012, The Digital Humanities Institute, Sheffield, UK.

Song, X., Ming, Z.-Y., Nie, L., Zhao, Y.-L., Chua, T.-S., 2016. Volunteerism Tendency Prediction via Harvesting Multiple Social Networks. *ACM Transactions on Information Systems* 34, 1–27. <https://doi.org/10.1145/2832907>

Sonnenwald, D.H., 2007. Scientific collaboration. *Annual Review of Information Science and Technology* 41, 643–681. <https://doi.org/10.1002/aris.2007.1440410121>

Spence, P., Brown, S., Meister, J.C., 2013. Introduction. *Literary and Linguistic Computing* 28, 491–492. <https://doi.org/10.1093/lrc/fqt062>

Spiro, L., 2012. “This Is Why We Fight”: Defining the Values of the Digital Humanities, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. University of Minnesota Press, pp. 16–35.

Spiro, L., 2011. Knowing and doing: Understanding the digital humanities curriculum, in: *Digital Humanities*. Stanford University Library, Stanford, pp. 232–233.

Spiro, L., 2009. Collaborative Authorship in the Humanities. *Digital Scholarship in the Humanities* 21.

Stanford University, 2019. What are the humanities? [WWW Document]. Stanford Humanities. URL <http://shc.stanford.edu/what-are-the-humanities> (accessed 12.24.19).

Stephan, P.E., 1996. The Economics of Science. *Journal of Economic Literature* 34, 1199–1235.

Stewart, B., 2016. Collapsed publics: Orality, literacy, and vulnerability in academic Twitter. *Journal of Applied Social Theory* 1, 61–86.

Stiegler, B., 2012. Call for digital studies [WWW Document]. *Digital Studies*. URL <https://digital-studies.org/wp/call-for-digital-studies/> (accessed 10.4.18).

Stock, W.G., Stock, M., 2013. *Handbook of information science*. De Gruyter Saur, Berlin ; Boston.

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>

Stutsman, S., 2013. Digital Humanities vs. New Media: A Matter of Pedagogy [WWW Document]. HASTAC. URL <https://www.hastac.org/blogs/stacistutsman/2013/11/17/digital-humanities-vs-new-media-matter-pedagogy> (accessed 6.26.19).

Suárez, J.L., 2010. Digital Humanities in Spanish? RED HD EN TRADUCCIÓN. URL <http://dayofdh2014.matrix.msu.edu/redhd/2014/04/08/digital-humanities-in-spanish/> (accessed 4.15.19).

Sugimoto, C.R., Ni, C., West, J.D., Larivière, V., 2015. The Academic Advantage: Gender Disparities in Patenting. *PLOS ONE* 10, e0128000. <https://doi.org/10.1371/journal.pone.0128000>

Sugimoto, C.R., Work, S., Larivière, V., Haustein, S., 2017. Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology* 68, 2037–2062. <https://doi.org/10.1002/asi.23833>

Suh, B., Hong, L., Pirolli, P., Chi, E.H., 2010. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network, in: 2010 IEEE Second International Conference on Social Computing. Presented at the 2010 IEEE Second International Conference on Social Computing (SocialCom), IEEE, Minneapolis, MN, USA, pp. 177–184. <https://doi.org/10.1109/SocialCom.2010.33>

Sula, C.A., Hill, H., 2017. The Early History of Digital Humanities. DH2017. URL <http://chrisalensula.org/the-early-history-of-digital-humanities/> (accessed 7.1.19).

Sutherland, J., 2006. The ideas interview: Franco Moretti [WWW Document]. The Guardian. URL <https://www.theguardian.com/books/2006/jan/09/highereducation.academicexperts> (accessed 6.23.19).

Svensson, P., 2016. Three Premises of Big Digital Humanities, in: Big Digital Humanities, Imagining a Meeting Place for the Humanities and the Digital. University of Michigan Press, pp. 82–130. <https://doi.org/10.2307/j.ctv65sx0t.7>

Svensson, P., 2012. Beyond the Big Tent, in: Gold, M.K. (Ed.), Debates in the Digital Humanities. University of Minnesota Press, pp. 36–49. <https://doi.org/10.5749/minnesota/9780816677948.003.0004>

Svensson, P., 2010. The Landscape of Digital Humanities 4.

Svensson, P., 2009. Humanities Computing as Digital Humanities 3.

Talja, S., Maula, H., 2003. Reasons for the use and non-use of electronic journals and databases: A domain analytic study in four scholarly disciplines. *Journal of Documentation* 59, 673–691. <https://doi.org/10.1108/00220410310506312>

Tambovtsev, Y.A., 1987. The Value of the Confidence Interval of the Consonant-Vowel Ratio as an Indicator of the Type of Linguistic Material. *Literary and Linguistic Computing* 2, 120–124. <https://doi.org/10.1093/lrc/2.2.120>

Tang, M.-C., Cheng, Y.-J., Chen, K., Hsiang, J., 2015. A Longitudinal Analysis of Knowledge Integration in Digital Humanities Using Co-citation Analysis, in: DH2015. Presented at the Digital Humanities 2015, Sydney, Australia.

Tang, M.-C., Cheng, Y.J., Chen, K.H., 2017. A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics* 113, 985–1008. <https://doi.org/10.1007/s11192-017-2496-6>

Taporwiki, 2011. How do you define humanities computing/digital humanities? [WWW Document]. URL http://www.artsrn.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing/_Digital_Humanities%3F

Taylor, R., 1990. Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography* 6, 35–39. <https://doi.org/10.1177/875647939000600106>

Teixeira, A.A.C., 2011. Mapping the (in)visible college(s) in the field of entrepreneurship. *Scientometrics* 89, 1–36. <https://doi.org/10.1007/s11192-011-0445-3>

Tello, J.C., 2017. Participants at the DH17 Conference by Country and Continent. CLIGS. URL <https://cligs.hypotheses.org/596> (accessed 10.18.18).

Tello, J.C., 2016. DHd 2016: countries, cities and institutions of the speakers. CLIGS. URL <https://cligs.hypotheses.org/431> (accessed 10.13.18).

Tello, J.C., 2015a. Visualizing information about HDH2015 & EADH Day. CLIGS. URL <https://cligs.hypotheses.org/261> (accessed 10.13.18).

Tello, J.C., 2015b. HDH 2015 and each Day: some impressions [WWW Document]. CLIGS. URL <https://cligs.hypotheses.org/257> (accessed 6.10.19).

Tennant, J.P., Penders, B., Ross-Hellauer, T., Marušić, A., Squazzoni, F., Mackay, A.W., Madan, C.R., Shaw, D.M., Alam, S., Mehmani, B., Graziotin, D., Nicholas, D., 2019. Boon, bias or bane? The potential influence of reviewer recommendations on editorial decision-making. *ESE* 45. <https://doi.org/10.20316/ESE.2019.45.18013>

Terras, M., 2016. A Decade in Digital Humanities. *Journal of Siberian Federal University. Humanities & Social Sciences* 9, 1637–1650. <https://doi.org/10.17516/1997-1370-2016-9-7-1637-1650>

Terras, M., 2012a. Infographic: Quantifying Digital Humanities. UCL DH Blog. URL <http://blogs.ucl.ac.uk/dh/2012/01/20/infographic-quantifying-digital-humanities/> (accessed 5.6.16).

Terras, M., 2012b. Image processing in the digital humanities, in: *Digital Humanities in Practice*. United Kingdom: Facet Publishing.

Terras, M., 2010. The Digital Classicist: Disciplinary Focus and Interdisciplinary Vision, in: Bodard, G., Mahony, S. (Eds.), *Digital Research in the Study of Classical Antiquity*. Ashgate.

Terras, M., 2006. Disciplined: Using Educational Studies to Analyse ‘Humanities Computing.’ *Lit Linguist Computing* 21, 229–246. <https://doi.org/10.1093/lit/fql022>

Terras, M., Nyhan, J., 2016. Father Busa's Female Punch Card Operatives, in: Debates in the Digital Humanities: 2016. University of Minnesota Press, Minneapolis London.

Terras, M.M., Nyhan, J., Vanhoutte, E. (Eds.), 2013. Defining Digital Humanities: A Reader. Ashgate Publishing Limited ; Ashgate Publishing Company, Farnham, Surrey, England : Burlington, VT.

Thaller, M., 2012. Controversies around the Digital Humanities: An Agenda. *Hist. Soc. Res.* 37, 7–23.

Tharenou, P., 1999. Gender differences in advancing to the top. *International Journal of Management Reviews* 1, 111–132. <https://doi.org/10.1111/1468-2370.00008>

The Alan Turing Institute, 2018. Data science and digital humanities [WWW Document]. The Alan Turing Institute. URL <https://www.turing.ac.uk/research/interest-groups/data-science-and-digital-humanities> (accessed 10.7.18).

The Editor, 1993. Back Matter. *Lit Linguist Computing* 8, 58–58.

Thelwall, M., Haustein, S., Larivière, V., Sugimoto, C.R., 2013. Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE* 8, e64841. <https://doi.org/10.1371/journal.pone.0064841>

Todres, Z.V., Todres, Z.V., 2009. Ion-radical organic chemistry: principles and applications, 2nd ed. ed. CRC Press, Boca Raton.

Tomlinson, J., 2009. Globalization and culture, Reprinted. ed. Polity Press, Cambridge.

Trajtenberg, M., 1990. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The RAND Journal of Economics* 21, 172. <https://doi.org/10.2307/2555502>

Tu, P.P.-N., 2011. A study of influential authors, works and research network of consumer behavior research. *African Journal of Business Management* 5, 9838–9854.

Tuire, P., Erno, L., 2001. Exploring Invisible Scientific Communities: Studying Networking Relations within An Educational Research Community: A Finnish Case. *Higher Education* 42, 493–513. <https://doi.org/10.1023/A:1012242009758>

Türker, İ., Sulak, E., 2017. A multilayer network analysis of hashtags in twitter via co-occurrence and semantic links. *International Journal of Modern Physics B* 1850029. <https://doi.org/10.1142/S0217979218500297>

Twitter, 2020a. How to receive recommendations from Twitter [WWW Document]. Twitter Help. URL <https://help.twitter.com/en/managing-your-account/how-to-receive-twitter-recommendations> (accessed 5.29.20).

Twitter, 2020b. About Twitter's account suggestions [WWW Document]. URL <https://help.twitter.com/en/using-twitter/account-suggestions> (accessed 5.29.20).

UK Gov, 2017. Graduates in the UK labour market - Office for National Statistics [WWW Document]. URL <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/graduatesintheuklabourmarket/2017#steady>

increase-in-the-number-of-graduates-in-the-uk-over-the-past-decade
(accessed 2.24.20).

Universities UK, 2018. Higher education research in facts and figures [WWW Document]. URL <https://www.universitiesuk.ac.uk/research-facts-and-figures> (accessed 2.24.20).

UNT, 2017. Registration for Digital Frontiers 2017 Now Open! [WWW Document]. Digital Scholarship. URL <https://blogs.library.unt.edu/digital-scholarship/2017/08/07/registration-for-digital-frontiers-2017-now-open/> (accessed 4.4.20).

US Census Bureau, 2018. Educational Attainment in the United States: 2018 [WWW Document]. The United States Census Bureau. URL <https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html> (accessed 2.24.20).

US National Science Board, 2020. Science and Engineering Indicators 2020 - Publications Output: U.S. Trends and International Comparisons [WWW Document]. National Science Foundation | National Science Board. URL <https://ncses.nsf.gov/pubs/nsb20206/> (accessed 2.3.20).

US National Science Board, 2010. Science and Engineering Indicators 2010 [WWW Document]. National Science Foundation | National Science Board. URL https://www.researchgate.net/publication/273929338_Science_and_Engineering_Indicators-1989 (accessed 2.3.20).

US National Science Board, 2000. Science and Engineering Indicators 2000 [WWW Document]. National Science Foundation | National Science Board. URL https://www.researchgate.net/publication/273929338_Science_and_Engineering_Indicators-1989 (accessed 2.3.20).

Usdiken, B., Pasadeos, Y., 1995. Organizational Analysis in North America and Europe: A Comparison of Co-citation Networks. *Organization Studies* 16, 503–526. <https://doi.org/10.1177/017084069501600306>

Uszkoreit, H., 2000. What is Computational Linguistics? [WWW Document]. CL Intro Text. URL http://www.coli.uni-saarland.de/~hansu/what_is_cl.html (accessed 6.25.19).

Uzuner, S., 2008. Multilingual scholars' participation in core/global academic communities: A literature review. *Journal of English for Academic Purposes* 7, 250–263. <https://doi.org/10.1016/j.jeap.2008.10.007>

van Arensbergen, P., van der Weijden, I., van den Besselaar, P., 2012. Gender differences in scientific productivity: a persisting phenomenon? *Scientometrics* 93, 857–868. <https://doi.org/10.1007/s11192-012-0712-y>

van Eck, N.J., 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>

van Eck, N.J., Waltman, L., 2007. VOS: A New Method for Visualizing Similarities Between Objects, in: Decker, R., Lenz, H.-J. (Eds.), *Advances in Data Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 299–306.

van Eck, N.J., Waltman, L., Noyons, E.C.M., Buter, R.K., 2010. Automatic term identification for bibliometric mapping. *Scientometrics* 82, 581–596. <https://doi.org/10.1007/s11192-010-0173-0>

van Lent, M., Overbeke, J., Out, H.J., 2014. Role of Editorial and Peer Review Processes in Publication Bias: Analysis of Drug Trials Submitted to Eight Medical Journals. *PLoS ONE* 9, e104846. <https://doi.org/10.1371/journal.pone.0104846>

Van Noorden, R., 2014. Online collaboration: Scientists and the social network. *Nature* 512, 126–129. <https://doi.org/10.1038/512126a>

Vandenbroucke, J.P., 1989. On not being born a native speaker of English. *BMJ: British Medical Journal* 298, 1461–1462.

Vanhoutte, E., 2013. The gates of hell: History and definition of digital/humanities/computing, in: *Defining Digital Humanities: A Reader*. Ashgate Publishing, Surrey/Burlington, pp. 119–158.

Veletsianos, G., 2012. Higher education scholars' participation and practices on Twitter: Scholars' participation and practices on Twitter. *Journal of Computer Assisted Learning* 28, 336–349. <https://doi.org/10.1111/j.1365-2729.2011.00449.x>

Veletsianos, G., Kimmons, R., 2016. Scholars in an increasingly open and digital world: How do education professors and students use Twitter? *The Internet and Higher Education* 30, 1–10. <https://doi.org/10.1016/j.iheduc.2016.02.002>

Verspagen, B., Werker, C., 2003. The invisible college of the economics of innovation and technological change. *Estudios de Economía Aplicada* 21, 393–419.

Vince, G., 2018. This is why good people turn bad online [WWW Document]. *The Independent*. URL https://www.independent.co.uk/news/long_reads/online-trolls-digital-societies-science-mary-beard-twitter-facebook-hate-speech-a8279596.html (accessed 4.17.20).

Vogel, R., 2012. The Visible Colleges of Management and Organization Studies: A Bibliometric Analysis of Academic Journals. *Organization Studies* 33, 1015–1043. <https://doi.org/10.1177/0170840612448028>

Wagner, C.S., 2008. The new invisible college: science for development. *Brookings Institution Press*, Washington, D.C.

Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., Rafols, I., Börner, K., 2011. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* 5, 14–26. <https://doi.org/10.1016/j.joi.2010.06.004>

Waltman, L., van Eck, N.J., Noyons, E.C.M., 2010. A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics* 4, 629–635. <https://doi.org/10.1016/j.joi.2010.07.002>

Wang, D., Cheng, H., Wang, P., Huang, X., Jian, G., 2017. Zipf's Law in Passwords. *IEEE Transactions on Information Forensics and Security* 12, 2776–2791. <https://doi.org/10.1109/TIFS.2017.2721359>

Wang, J., Zhao, W.X., He, Y., Li, X., 2014. Infer User Interests via Link Structure Regularization. *ACM Transactions on Intelligent Systems and Technology* 5, 1–22. <https://doi.org/10.1145/2499380>

Wang, M.C., 2016. Digitization and the Depoliticization of Museum Access The Situation at the National Palace Museum. *Advances in Computer Science : an International Journal* 5, 40–46.

Wang, R., Liu, W., Gao, S., 2016. Hashtags and information virality in networked social movement. *Online Information Review*. <https://doi.org/10.1108/OIR-12-2015-0378>

Wang, X., Inaba, M., 2009a. Analyzing Structures and Evolution of Digital Humanities Based on Correspondence Analysis and Co-word Analysis. *Art research* 9, 123–134.

Wang, X., Inaba, M., 2009b. Structures and Evolution of Digital Humanities: An Empirical Research based on Correspondence Analysis and Co-word Analysis.

Wang, X., Xu, S., Wang, Z., Peng, L., Wang, C., 2013. International scientific collaboration of China: collaborating countries, institutions and individuals. *Scientometrics* 95, 885–894. <https://doi.org/10.1007/s11192-012-0877-4>

Wardrip-Fruin, N., Montfort, N. (Eds.), 2003. *The New Media Reader*. MIT Press, Cambridge, Mass.

Warwick, C., 2015. Building Theories or Theories of Building? A Tension at the Heart of Digital Humanities, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A New Companion to Digital Humanities*. John Wiley & Sons, Ltd, Chichester, UK, pp. 538–552. <https://doi.org/10.1002/9781118680605.ch37>

Warwick, C., 2012. Institutional models for digital humanities, in: *Digital Humanities in Practice*. pp. 193–216.

Warwick, C., Terras, M., Nyhan, J., 2012. *Digital Humanities in Practice*. United Kingdom: Facet Publishing.

Wasburn, M.H., 2007. Mentoring women faculty: an instrumental case study of strategic collaboration. *Mentoring & Tutoring: Partnership in Learning* 15, 57–72. <https://doi.org/10.1080/13611260601037389>

Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. *Nature* 393, 440–442. <https://doi.org/10.1038/30918>

Web of Science, 2017. InCites Journal Citation Reports [WWW Document]. Clarivate Analytics. URL <http://jcr.incites.thomsonreuters.com/JCRHomePageAction.action?>

Weingart, S.B., 2018. dh quantified: A Review of Quantitative Analyses of the Digital Humanities. the scottbot irregular: data are everywhen. URL <http://scottbot.net/dh-quantified/>

Weingart, S.B., 2016a. Representation at Digital Humanities Conferences (2000–2015).

Weingart, S.B., 2016b. Submissions to DH2017 (pt. 1) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/?p> (accessed 10.13.18).

Weingart, S.B., 2015a. Submissions to DH2016 (pt. 1) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/submissions-to-dh2016-pt-1/> (accessed 10.13.18).

Weingart, S.B., 2015b. Acceptances to Digital Humanities 2015 (part 1) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/acceptances-to-digital-humanities-2015-part-1/> (accessed 10.17.18).

Weingart, S.B., 2015c. Acceptances to Digital Humanities 2015 (part 2) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/acceptances-to-digital-humanities-2015-part-2/> (accessed 10.17.18).

Weingart, S.B., 2015d. Acceptances to Digital Humanities 2015 (part 3) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/acceptances-to-digital-humanities-2015-part-3/> (accessed 10.17.18).

Weingart, S.B., 2015e. Acceptances to Digital Humanities 2015 (part 4) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/acceptances-to-digital-humanities-2015-part-4/> (accessed 10.17.18).

Weingart, S.B., 2014a. Submissions to Digital Humanities 2015 (pt. 1) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/submissions-to-digital-humanities-2015-pt-1/> (accessed 10.13.18).

Weingart, S.B., 2014b. Acceptances to Digital Humanities 2014 (part 1) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/acceptances-to-digital-humanities-2014-part-1/> (accessed 10.17.18).

Weingart, S.B., 2014c. Submissions to Digital Humanities 2015 (pt. 3) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/submissions-to-digital-humanities-2014-pt-3/> (accessed 10.17.18).

Weingart, S.B., 2013a. Submissions to Digital Humanities 2014 [WWW Document]. the scottbot irregular. URL <http://scottbot.net/submissions-to-digital-humanities-2014/> (accessed 10.13.18).

Weingart, S.B., 2013b. Acceptances to Digital Humanities 2013 (part 1) [WWW Document]. the scottbot irregular. URL <http://scottbot.net/analyzing-acceptances-to-digital-humanities-2013-part-1/> (accessed 10.17.18).

Weingart, S.B., 2012. Analyzing submissions to Digital Humanities 2013 [WWW Document]. the scottbot irregular. URL <http://scottbot.net/digital-humanities-2013-submission-analysis/> (accessed 10.13.18).

Weingart, S.B., Eichmann-Kalwara, N., 2017. What's Under the Big Tent?: A Study of ADHO Conference Abstracts. *Digital Studies/Le champ numérique* 7. <https://doi.org/10.16995/dscn.284>

Weingart, S.B., Eichmann-Kalwara, N., Jorgensen, J., 2016. Gender & Centrality DH 2000-2016.

Weller, M., 2011. The Digital Scholar: How Technology is Changing Academic Practice. Bloomsbury Publishing PLC. <https://doi.org/10.5040/9781849666275>

Wernimont, J., 2018. Women in DH [WWW Document]. Google Docs. URL <https://docs.google.com/spreadsheets/d/1pPscJX7I7Vvuc4YIBbd38nbWgIjp-0Fil5yZ1sxG6Vk/edit#gid=0> (accessed 10.17.18).

Wernimont, J., 2015. Introduction to Feminisms and DH special issue. DHQ 009.

Wernimont, J., 2013. Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives. DHQ: Digital Humanities Quarterly 7.

White, H.D., 1990. Author co-citation analysis: Overview and defense. *Scholarly communication and bibliometrics* 84, 106.

White, H.D., Griffith, B.C., 1981. Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science* 32, 163–171. <https://doi.org/10.1002/asi.4630320302>

White, H.D., McCain, K.W., 1998. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American society for information science* 49, 327–355.

Williams, S.A., Terras, M.M., Warwick, C., 2013. What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation* 69, 384–410. <https://doi.org/10.1108/JD-03-2012-0027>

Williamson, V., 2016. On the ethics of crowd sourced research. *PS: Political Science & Politics* 49, 77–81.

Wisbey, R., 1962. Concordance Making by Electronic Computer: Some Experiences with the “Wiener Genesis.” *The Modern Language Review* 57, 161–172. <https://doi.org/10.2307/3720960>

Wittgenstein, L., 1953. Philosophical investigations: the German text, with a revised English translation, 3rd ed. ed. Blackwell Pub, Malden, MA.

Witting, A., 2018. A Sociometric Analysis of the Policy Learning Community. <https://doi.org/10.13140/rg.2.2.29396.50561>

Wolff, G.W., 1974. A Scalogram analysis of the Kansas-Nebraska bill of 1854 and related roll-calls in the house of representatives. *Comput Hum* 8, 71–83. <https://doi.org/10.1007/BF02530740>

Wong, B., 2011. Color blindness. *Nature Methods* 8, 441–441. <https://doi.org/10.1038/nmeth.1618>

Wortham, J., 2009. Bit.ly Eclipses TinyURL on Twitter. *Bits Blog*. URL <https://bits.blogs.nytimes.com/2009/05/07/bitly-eclipses-tinyurl-on-twitter/> (accessed 5.18.20).

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. *Science* 316, 1036–1039. <https://doi.org/10.1126/science.1136099>

WWW2018, 2018. Crowdsourcing and Human Computation for the Web. URL <https://www2018.thewebconf.org/> (accessed 10.7.18).

Yan, E., Ding, Y., 2009. Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology* 60, 2107–2118. <https://doi.org/10.1002/asi.21128>

Yan, X.-S., 2011. Information Science: Its Past, Present and Future. *Information* 2, 510–527. <https://doi.org/10.3390/info2030510>

Yang, G., 2016. Narrative Agency in Hashtag Activism: The Case of #BlackLivesMatter. *MaC* 4, 13. <https://doi.org/10.17645/mac.v4i4.692>

Yin, D., Hong, L., Xue, Z., Davison, B.D., 2011. Temporal Dynamics of User Interests in Tagging Systems, in: Proceedings of the 25th AAAI Conference on Artificial Intelligence. Presented at the AAAI Conference on Artificial Intelligence, pp. 1279–1285.

Zaagsma, G., 2013. On Digital History. *BMGN - Low Countries Historical Review* 128, 3. <https://doi.org/10.18352/bmgn-lchr.9344>

Zhao, D., Logan, E., 2002. Citation analysis using scientific publications on the Web as data source: A case study in the XML research area. *Scientometrics* 54, 449–472. <https://doi.org/10.1023/A:1016090601710>

Zhao, D., Strotmann, A., 2015. Analysis and Visualization of Citation Networks.

Zhao, D., Strotmann, A., 2011. Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics* 87, 115–131. <https://doi.org/10.1007/s11192-010-0317-2>

Zhou, D., Huang, J., Schölkopf, B., 2005. Learning from labeled and unlabeled data on a directed graph. ACM Press, pp. 1036–1043. <https://doi.org/10.1145/1102351.1102482>

Zhu, B., Nie, H., 2017. 互动与共生:数字人文与史学研究——第二届“北京大学数字人文论坛”综述. *大学图书馆学报* 2017, 18–22. <https://doi.org/10.16603/j.issn1002-1027.2017.04.003>

Zhu, B., Nie, H., 2016. 跨界与融合: 全球视野下的数字人文——首届北京大学“数字人文论坛”会议综述. *大学图书馆学报* 2016, 16–21.

Zitt, M., Bassecoulard, E., Okubo, Y., 2000. Shadows of the Past in International Cooperation: Collaboration Profiles of the Top Five Producers of Science. *Scientometrics* 47, 627–657. <https://doi.org/10.1023/A:1005632319799>

Zorich, D., 2008. A survey of digital humanities centers in the United States. Council on Library and Information Resources, Washington, D.C.

Zuccala, A., 2006. Modeling the Invisible College. *J. Am. Soc. Inf. Sci.* 57, 152–168. <https://doi.org/10.1002/asi.20256>

Zuccala, A., 2004. Revisiting the Invisible College: A Case Study of the Intellectual Structure and Social Process of Singularity Theory Research in Mathematics. University of Toronto.

Zuccala, A., van den Besselaar, P., 2009. Mapping Review Networks: Exploring Research Community Roles and Contributions. *Scientometrics* 81, 111–122. <https://doi.org/10.1007/s11192-008-2136-2>

