

Joint and multi-authored publication patterns in the Digital Humanities

Julianne Nyhan and Oliver Duke-Williams

Department of Information Studies, UCL

Abstract

The stereotype of the multi-authored Digital Humanities paper is well known but has not, until now, been empirically investigated. Here we present the results of a statistical analysis of collaborative publishing patterns in *Computers and the Humanities* (CHum) (1966–2004); *Literary and Linguistic Computing* (LLC) (1986–2011); and, as a control, the *Annals of the Association of American Geographers* (AAAG) (1966–2013) in order to take a first step towards investigating concepts of ‘collaboration’ in Digital Humanities. We demonstrate that in two core Digital Humanities journals, CHum and LLC, single-authored papers predominate. In AAAG, single-authored papers are also predominant. In regard to multi-authored papers the statistically significant increases are more wide-ranging in AAAG than in either LLC or CHum, with increases in all forms of multi-authorship. The author connectivity scores show that in CHum, LLC, and AAAG, there is a relatively small cohort 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.

Correspondence:

Julianne Nyhan, Department of Information Studies, Foster Court, University College London, Gower Street, London, WC1E 6BT.

Email:

j.nyhan@ucl.ac.uk

1 Introduction

The stereotype of the Humanities lone scholar is well known, even if it is increasingly recognized as impoverished (see, for example, Bulger *et al.*, 2011; McGrath, 2011: 2). The stereotype of the collaborative Digital Humanities (DH) scholar is also well known (see, for example, Fitzpatrick, 2011; Moulin *et al.*, 2011; Deegan and McCarty, 2012; MLA, 2012). This collaborative nature is frequently presented as something that separates the Digital Humanities from the traditional Humanities, for example:

‘Traditional humanities scholarship rewards the solitary endeavour (such as the single-authored monograph) and looks askance at collaboration (e.g. edited volumes), but many digital humanities projects are often collaborative in nature. This translates to an ethos of sharing and collegiality in these environments,

but the multi-author aspect of these digital projects may cause problems during evaluation’ (Koh, 2012).

Willard McCarty has written that ‘Collaboration is a problematic and should be a contested term’ (McCarty, 2012: 2). Indeed, despite the seeming ubiquity of collaboration in Digital Humanities, it is interesting to note that relatively few studies have been undertaken into the kinds of collaborative activities that take place in DH, and how they can be identified and measured. In the quotation from Koh, above, the implication is that multi-authorship is an indicator of collaboration and this is a reasonable inference to make. Similar assumptions can also be noticed in, *inter alia*, Borgman (2009) who writes, citing Cronin (2005), ‘The size of collaborations is increasing in all fields, as measured by the number of co-authors on papers, and at the fastest rate in the sciences’.

However, literature from the domain of information science emphasizes that the issue may be more complicated than this and that collaboration and co-authorship do not necessarily have a one-to-one or collocative relationship. It is widely known that the order that authors are listed in and how this relates to the relative contributions of each can vary from field to field. Indeed, by 1958, Smith had emphasized the need to gather information from listed authors about their input to a given paper, an idea further developed and refined by Subramanyam (1983), who emphasized the difficulties of capturing such information. Katz and Martin (1997) have demonstrated that co-authorship can be viewed as but one indicator of collaboration, and one that is valid only where co-authors are actually specified. Further to this, Laudel (2002) has emphasized the kinds of collaboration that can remain hidden when the actors a contributing individual's name is not listed as an author on an academic paper. It is not just definitions of collaboration that vary, definitions of authorship do too. Bošnjak, L. and Marušić, A., (2012), for example, reviewed

Prevalence of authorship statements, their specificity and tone, and contributions required for authorship were assessed in 185 journals from *Science Citation Index* (SCI) and *Social Science Citation Index* (SSCI), 260 journals from *Arts & Humanities Citation Index* (A&HCI) and 651 codes of ethics from professional organizations from the online database of the Center for the Study of Ethics in the Profession, USA. In SCI, 53% of the top-ranked journals had an authorship statement, compared with 32% in SSCI. In a random sample of A&HCI-indexed journals, only 6% of the journals addressed authorship.

Looking back to the DH context, core journals such as *Literary and Linguistic Computing* (LLC) and *Digital Humanities Quarterly* do not, to the best of our knowledge, include authorship statements. Given the wide range of disciplines that publish under the umbrella term 'DH' it may well be that a suitable definition is impossible to reach. Nevertheless, information about definitions of authorship can be gleaned from the publications

of various research groups. For example, in addition to specifying co-authorship of presentations, the Implementing New Knowledge Environments (NKE) project specifies that all members of the INKE research team:

...receive named co-authorship credit on presentations and publications that make direct use of research in which they took an active, as opposed to passive, role (i.e. research to which the individual made a unique and discernable contribution with a substantial effect on the knowledge generated); otherwise, receive indirect credit via the INKE corporate authorship convention (Siemens *et al.*, 2009).

So too the project charter written by Ruecker and Radzikowska (2008) notes that

For presentations or papers where this work is the main topic, all team members who worked directly on this subproject should be co-authors. Any member can elect at any time not to be listed, but may not veto publication... For presentations or papers that spin off from this work, only those members directly involved need to be listed as co-authors. The others should be mentioned if possible in the acknowledgments, credits, or article citations (2008).

Presumably if such practices were 'par for the course' in Digital Humanities as a whole, they would not need to be articulated.

Considering this literature and its applications it seems that the interconnections in DH between collaboration and multi-authored papers may be more complex than is often assumed. In order to take a first step towards investigating such a wide-ranging concept as 'collaboration' we set out to gain a better understanding of how publication patterns in the context of academic journals have changed since the first journal of the field, *Computers and the Humanities*, was established in 1966. With the above caveats in mind we asked the following: has joint publication become more or less common or have the proportions of jointly published articles remained the same? How do such patterns compare with other disciplines of the Humanities? Is there

any mismatch between the collaborative way that the field describes itself and what we find when we examine the evidence of publication patterns and practices in its key journals?

To the best of our knowledge, the empirical evidence of publication practices of Digital Humanities scholars has not, until our research, been systematically investigated.

2 Research Context

Looking to the Humanities more generally, and across the scientific domains, it is clear from the quantitative studies that have been carried out that with a few exceptions, co- or multi-authorship has been increasing across the disciplines since the post-World War II period. One of the first to notice this was Smith (1958) whose analysis of publications in *American Psychologist* published between 1946 and 1957 revealed an increase in co-authorship. Authors such as Cronin have also showed that co-authorship across disciplines has been increasing, with the largest increases in science (Cronin, 2005). More recently, Wuchty *et al.* (2007) studied 19.9 million papers and 2.1 million papers over some 50 years in the Institute for Scientific Information (ISI) Web of Science database. They found that

In sciences and engineering, 99.4% of the 171 subfields have seen increased teamwork. Meanwhile, 100% of the 54 subfields in the social sciences, 88.9% of the 27 subfields in the humanities and 100% of the 36 subfields in patenting have seen increased teamwork. . . . As reflected in Figure 1A, the humanities show lower growth rates in the fraction of publications done in teams, yet a tendency toward increased teamwork is still observed. (p. 1037)

This has also been borne out in studies of particular geographical or disciplinary areas. Changing publication patterns in the Humanities in Flanders and Belgium have been analysed by Engels *et al.* (2012) who found that in the period 2000–09 ‘The overall growth rate in number of publications is over 62.1%, but varies across disciplines between 7.5

and 172.9%. Publication output grew faster in the Social Sciences than in the Humanities’. In 2003, Kyvik found that in Norwegian Universities co-authorship has become more common but it is difficult to determine from the article to what degree this applies to the Humanities. Lariviere *et al.* (2006) used data from the CD-ROM versions of the *Science Citation Index*, *Social Sciences Citation Index* and the *Arts & Humanities Citation Index* from 1980 to 2002, to argue that

contrary to a widely held belief, researchers in the social sciences and the humanities do not form a homogeneous category. In fact, collaborative activities of researchers in the social sciences are more comparable to those of researchers in the [natural sciences and engineering] than in the humanities (2006).

A number of publications have reflected on the new collaborative possibilities that electronic publishing can offer and that digital humanities work can require (see, for example, Kelleher *et al.*, 2011; McGrath 2011). However, looking to studies of collaborative publication patterns in Digital Humanities, it is clear that relatively little has been done on this from a quantitative perspective. The main exception is Spiro (2011) and Spiro (2009):

I compared the number of collaboratively-written articles published between 2004 and 2008 in two well-respected quarterly journals, *American Literary History* (ALH) and *Literary and Linguistic Computing* (LLC). . . . So what did I find? Whereas 5 of 259 (1.93%) articles published in ALH—about one a year—feature two authors (none had more than two), 70 out of 145 (48.28%) of the articles published in LLC were written by two or more authors. Most (4 of 5, or 80%) of the ALH articles were written by scholars from multiple institutions, whereas 49% (34 of 70) of the LLC articles were. About 16% (11 of 70) of the LLC articles featured contributors from two or more countries, while none of the ALH articles did. . . . the initial results do suggest that collaborative authorship is more common in digital humanities.

To the best of our knowledge, the empirical, statistical evidence of publication practices of Digital Humanities scholars has not, until our research, been systematically investigated. In order to make a first contribution towards addressing this gap in the research literature, we focused our research on publication patterns in two leading Digital Humanities journals since 1966.

3 Methodology

Using Zotero we extracted bibliographical meta-data from *Computers and the Humanities* (Chum) (1966–2004); *Literary and Linguistic Computing* (LLC) (1986–2011); and, as a control (see below), the *Annals of the Association of American Geographers* (1966–2013).

In order to establish a comparative context for the trends in multi-authorship that we observed in DH, we included the *Annals of the Association of American Geographers* in our analysis.

We selected this journal because it is a respected Geography journal that attracts a range of research, including research with technical applications or methodologies, for example, GIS. As specified on their website:

The *Annals of the Association of American Geographers* publishes original, timely, and innovative peer-reviewed articles that advance knowledge in all facets of the discipline. These articles address significant research problems and issues, and are attuned to the sensibilities of a diverse scholarly audience . . . [with] articles in four major areas—Environmental Sciences; Methods, Models, and Geographic Information Science; Nature and Society; and People, Place, and Region. (AAAG, n.d.)

Thus, while having a solid humanities context, this journal also includes social science, computer and engineering science, and information science perspectives, this being in some way comparable with the range of researchers that both submit work to and read CHum and LLC. From a purely practical perspective, one of the authors of this article also has expertise in geography and so when issues

about, for example, differentiation of names arose (see further below) we were able to draw on our personal experience to make decisions in addition to carrying out web searches of institutional faculty profile pages etc.

Once harvested, all data were exported to Excel for initial viewing. As far as possible, the data were cleaned and regularized—e.g. a canonical form of personal names was chosen where slight differences existed such as E.G. Wills and Edward G. Wills. A look-up table was used in the database to map different variant spellings of a person's name to a single canonical version. This occurs quite commonly, and is more common for people who have published heavily over a long time: in the worst cases, there were up to five different spellings of a person's name. This table had to be prepared by hand, and this obviously risks both false pairings (two very similar names which are not, in fact, the same person) and omissions (two names which should map to the same person but are treated as separate people). Neither sort of error should affect the analysis of the frequency of papers with different numbers of authors, but might affect some of the other analyses

In all of the journals, there were a large number of papers that we wanted to exclude from analysis, as they are not peer reviewed, and even though it is less likely that these would be co-authored, there are enough of them to skew the data. These include reviews, letters to the editor, responses, errata, notes on conferences and seminars, etc. The data and metadata we harvested using Zotero did not do a good job of distinguishing between research papers and other content; though this may vary from journal to journal. Fortunately, there is an approach that can be used (though not with complete accuracy) to differentiate between publication types: identifying titles that have been used more than once, as these are generally formulaic titles such as 'book review'. In our SQL analysis, we read in details of papers and authors, followed by a separately prepared list of authors with identified aliases (alternative spellings etc). Following various preparatory SQL data processing steps, we identified all cases where two or more papers shared a common title. These were examined to see whether any such pairs (groupings) of papers were genuine (different

research papers, common title). This does occur, although it is rare in our data. In some cases, this may be because a subtitle has been omitted in our bibliographic data. All papers with common titles were then removed (apart from any identified exceptions). In the case of CHum, 198 of 1241 papers were moved (c. 16%) In the case of LLC, 134 of 1050 papers were removed (c. 13%)

The cleaned data were then imported into an SQL database and sorted into groups based on the number of authors. The annual observed frequencies of papers with n authors were then calculated. For each group, a linear regression was calculated in order to determine within a given journal whether the incidence of n -authored papers had changed over time.

For each journal, the data were also processed so that dual-authored papers could be analysed using a connectivity index (Bell *et al.*, 2002) to determine the extent to which the pool of authors contributing to a given journal were interconnected. A connectivity index was constructed both on a journal-wide basis and on a per-author basis, allowing the distribution of ‘well-connected’ authors to be compared within and between journals.

A migration connectivity index (Bell *et al.*, 2002) is a summary indicator that describes the amount of migration between zones, typically a set of subnational areas. It varies between possible values of 0.0, indicating that no migration between areas has occurred in the observed period, and 1.0 indicating that at least one person moved from each zone to all other zones in the system, and that all pairs of zones are thus *connected*. An analogous index can be constructed for linkages between collaborative authors, for a given journal.

Given n different authors who have published (whether as a sole author, or as part of a multiple author group) in a journal in a given period, the total number of possible linkages between authors is $n(n-1)$ —each author could, potentially, have collaborated with each other author. By analysing the bibliographic metadata, we can determine the total number of distinct collaborations m between different authors over the same period. To generate this total, we add one to m for each observed case of two authors collaborating, regardless of how many

papers they jointly published. An index i can thus be constructed:

$$i = \frac{m}{(n(n-1))/2}$$

that measures the connectedness of authors in the chosen journal. We divide the denominator by 2, as any observed co-publishing between authors is counted equally as a collaboration for both, halving the total number of potential connections.

For individual authors, a revised index can be constructed. For an individual author j we let m' be the total number of different authors with which j has co-published. The total number of possible authors with which this might have happened is $n-1$, and thus an individual index can be constructed:

$$i_j = \frac{m'}{n-1}$$

For i , constructed at the journal level, the index would have a value of 0 if no authors had co-published (i.e. all papers were singly authored) and a value of 1 if all authors had published at least one co-authored paper with all other authors. Similarly, for the individual index i_j , a value of 0 would indicate that all of that author’s publications were singly authored, whilst a value of 1 would indicate that they had published at least one paper with all other authors who had published in the journal in the observed period.

Clearly, these are theoretical constraints, and whilst a value of zero is feasible, a value of one is unlikely to be observed.

4 Results

4.1 Computers and the humanities (1966–2004): findings

In Figure 1, it is shown that the trend over time is for single authorship to predominate and hold steady. We note that there is very marked variation year-by-year in the number of single-authored papers: by selecting a particular subset of years rather than the full-time range, a rising or falling trend could be demonstrated if one so desired. There is a significant increase in dual- and triple-authored papers but the increases for four- and five-authored papers are not significant (see Table 1).

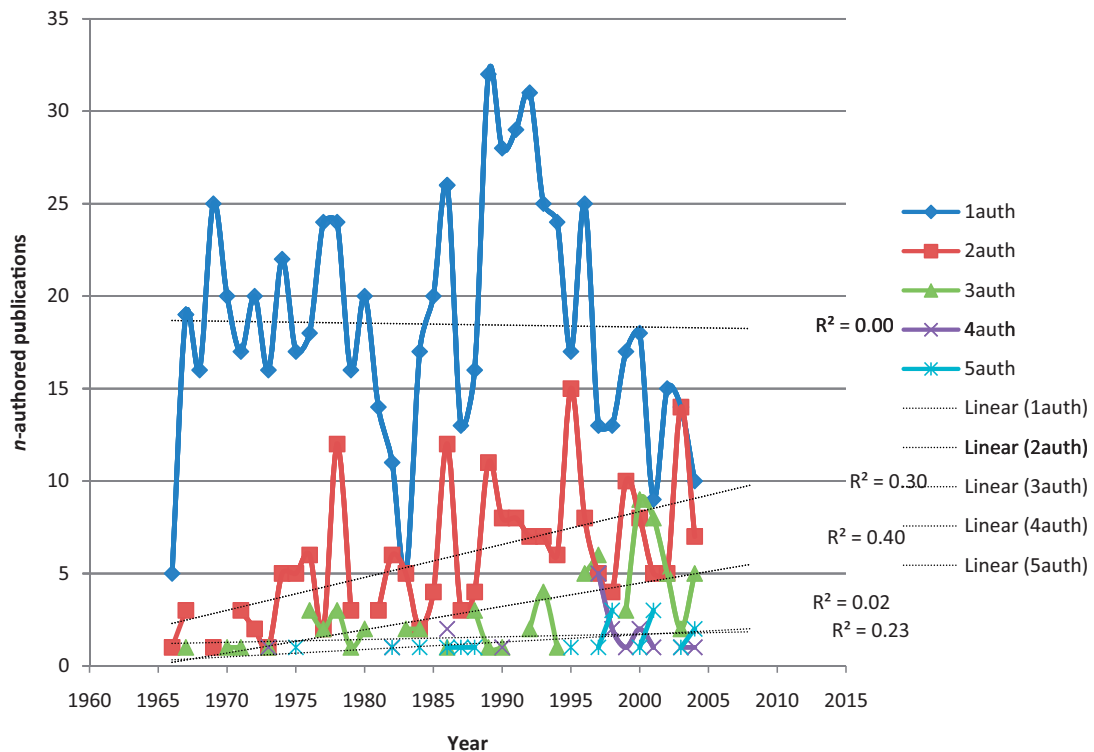


Fig. 1 CHum graph

Table 1 CHum correlations

Correlations		Year	Auth1	Auth2	Auth3	Auth4	Auth5
Year	Pearson correlation	1	−0.018	0.546**	0.634**	0.123	0.479
	Significance (two-tailed)		0.913	0.001	0.001	0.720	0.098
	N	39	39	36	26	11	13

**Correlation is significant at the 0.01 level (two-tailed).

4.2 Literary and linguistic computing: findings

In Figure 2, it is shown that in LLC (1986–2011), single-authored papers are predominant and though the trend over time shows a decrease in the number of single-authored papers, this decrease is not statistically significant. Again, there is strong fluctuation in the observed number of single-authored papers. A significant increase in triple-authored papers can be observed but this is not the case for

joint-authored or four- or five-authored papers (see Table 2)

4.3 Annals of the Association of American Geographers (1966–2013): findings

In Figure 3, it is shown that single-authored papers are predominant, and though the trend over time shows a slight increase, it is not statistically significant; statistically significant increases in

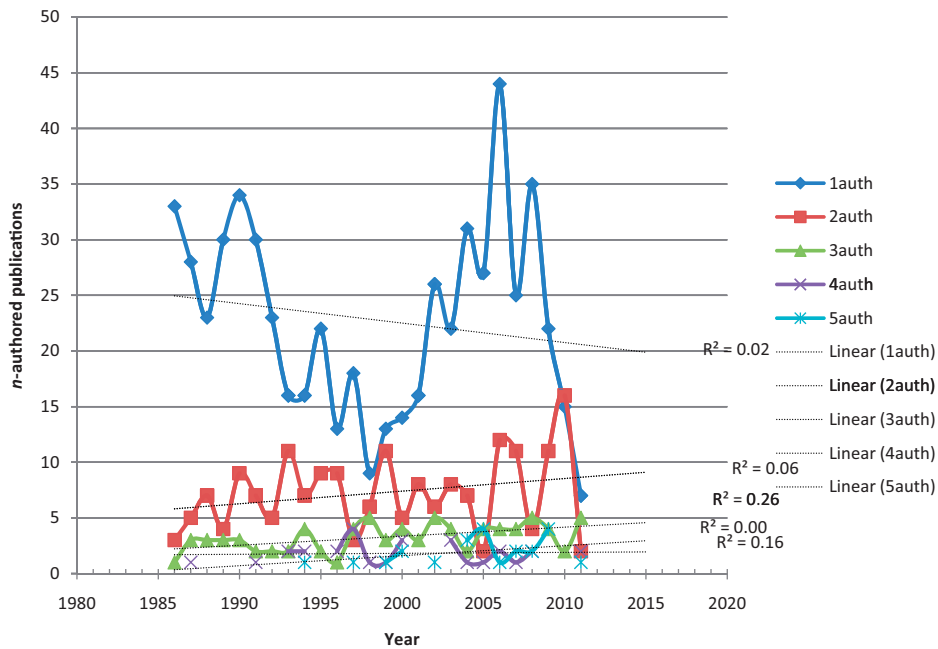


Fig. 2 LLC graph

Table 2 LLC correlations

Correlations		Year	Auth1	Auth2	Auth3	Auth4	Auth5
Year	Pearson correlation	1	−0.150	0.252	0.514**	0.063	0.399
	Significance (two-tailed)		0.465	0.214	0.007	0.817	0.198
	N	26	26	26	26	16	12

**Correlation is significant at the .01 level (two-tailed).

collaborative publishing patterns can be seen in all other sets of *n*-authored papers (see Table 3).

4.4 Connectivity index findings: CHum, LLC, and AAAG

The connectivity analysis showed a journal level index value of 0.0012 for LLC, and 0.00096 for CHum. Thus, co-publishing within the discipline (self-defined as ‘those who publish in the journal’) appears to be more strongly observed in LLC than in CHum. This might be an effect of the differing time frames over which the journals were analysed. A journal level index was also constructed for AAAG, with data over a similar length of time to CHum, and was found to be lower at 0.00057. Further analysis of a wider set of journals would

be beneficial to gain a greater understanding of the range of variation of this index in journals in different disciplines.

Of more immediate interest, author level connectivity scores were also constructed for the three journals. All three journals display similar characteristics: there is 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. We calculated an author level index for all authors, and then aggregated these to determine the number of authors by index value. Thus, for example, in LLC, we observed that 1 author had co-published with 13 other authors (giving an individual index value of 0.013), the next highest observations were of 1 author who had co-published with

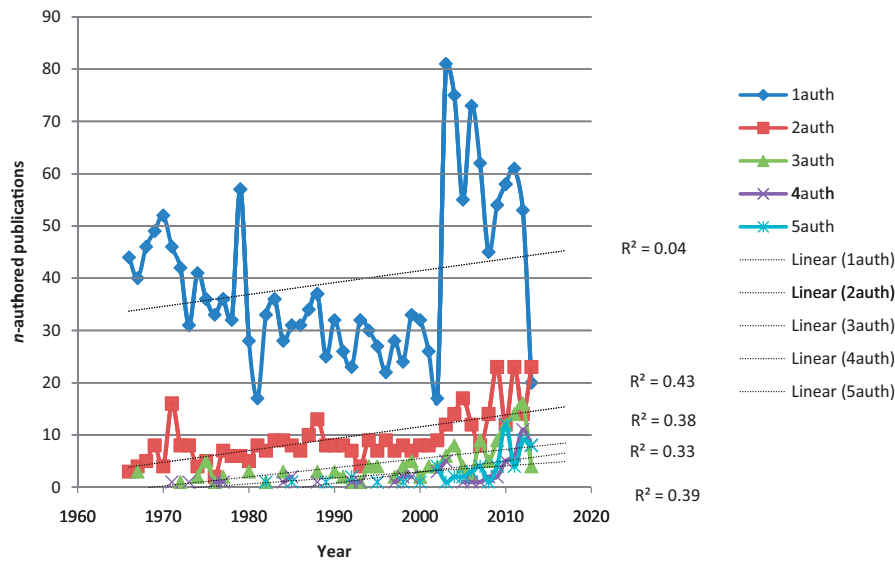


Fig. 3 AAAG graph

Table 3 AAAG correlations

Correlations		Year	Auth1	Auth2	Auth3	Auth4	Auth5
Year	Pearson correlation	1	0.210	0.653**	0.616**	0.575**	0.623**
	Significance (two-tailed)		0.153	0.000	0.000	0.003	0.004
	N	48	48	48	33	24	19

**Correlation is significant at the 0.01 level (2-tailed).

9 other authors, and then 3 authors who had each co-published with 8 other authors. For CHum, the highest observation was of one author who had co-published with 8 other authors, and for AAAG, we observed one author who had co-published with 13 other authors (although, given a wider overall pool of authors, this generated an individual index value of 0.005). Table 4 below, shows—for the three journals—the highest observations; the number of authors shown for each journal varies to accommodate all authors observed with a particular index value.

In Figure 4, it is demonstrated that the profiles show LLC to have both higher index values than is the case for CHum and AAAG for the most widely co-publishing authors, and also higher numbers of authors observed for other index values. LLC and Chum have a similar final observation, whereas AAAG has a longer tail with a larger number of

authors with a low index value, reflecting the fact that AAAG is diverse in the number of subdisciplines of geography that it covers, and also that geography in general is a much broader discipline than DH, and thus has a much wider pool of authors. However, that wider pool of authors does not include a subset that has any greater propensity to co-publish than is the case in LLC.

5 Summary and Analysis

In CHum we found that single-authored papers are predominant and the trend over time holds steady; there is a significant increase in dual- and triple-authored papers. In LLC (1986–2011), single-authored papers are predominant; the trend over time shows a significant increase in triple-authored

Table 4 Connectivity index for LLC, CHum, and AAAG

LLC			CHum			AAAG		
Author name ^a	Observed different co-authors	Index value	Author name	Observed different co-authors	Index value	Author name	Observed different co-authors	Index value
LLC_A	13	0.0133	CHUM_A	8	0.0067	AAAG_A	13	0.0052
LLC_B	9	0.0092	CHUM_B	7	0.0058	AAAG_B	12	0.0048
LLC_C	8	0.0082	CHUM_C	6	0.0050	AAAG_C	11	0.0044
LLC_D	8	0.0082	LLC_F	6	0.0050	AAAG_D	10	0.0040
LLC_E	8	0.0082	CHUM_D	6	0.0050	AAAG_E	9	0.0036
LLC_F ^b	7	0.0072	CHUM_E	5	0.0042	AAAG_F	8	0.0032
LLC_G	6	0.0061	CHUM_F	5	0.0042	AAAG_G	8	0.0032
LLC_H	6	0.0061	CHUM_G	5	0.0042	AAAG_H	8	0.0032
LLC_I	6	0.0061	CHUM_H	5	0.0042	AAAG_I	8	0.0032
			CHUM_I	5	0.0042	AAAG_J	8	0.0032
			CHUM_J	5	0.0042			

^aAuthor names by nature identify individuals, and have been masked here, pending ethical advice on whether they can be used.

^bOne author ('LLC_F') is common between the two subsets of authors shown for LLC and CHum; were these lists to be extended to show a greater number of authors, more overlaps would be apparent.

papers. Despite the stereotype of the DH researcher as one who engages in collaborative research and publishes multi-authored papers, our findings indicate that single authorship is predominant in the sample we took from some of the core journals of the field (CHum (1966–2004) and LLC (1986–2011)). Perhaps this should not come as a surprise given the 'humanities' aspect of DH and the role of single-authorship in the humanities. It is also important to emphasize that the predominance of single-authorship does not necessarily point to the absence (or even overestimation) of the role of collaboration in DH. These figures do, however, point to the necessity of reflecting further on the types of collaborative partnerships and activity that DH routinely acknowledges (or not) as well as the varied ways that it acknowledges such partnerships.

Considering the comparative context in AAAG (1966–2013), similar to CHum and LLC, single-authored papers are predominant. However, it is notable that in contrast to CHum and LLC, in AAAG the increase in multi-authored papers is statistically significant in all of its forms (joint-authored through to five-authored papers).

The literature review above showed that an increase in multi-authored papers can be noticed in most disciplines since the post-World War II

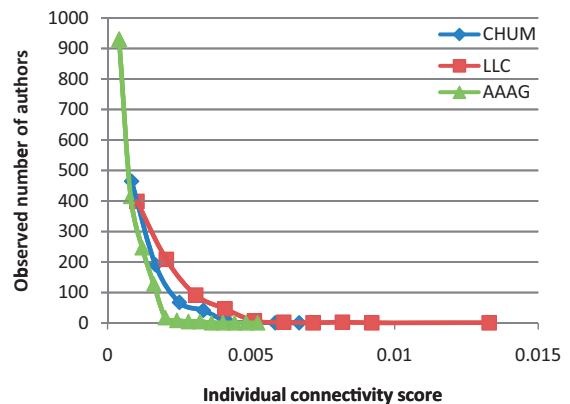


Fig. 4 The total number of authors observed in three journals who had a common connectivity index value

period. So it seems that, based on the CHum and LLC data, digital humanities is not at all unusual in showing statistically significant increases in multi-authored papers. Perhaps more surprising is that a journal such as AAAG shows more significant increases in multi-authorship. Whatever the reasons for this (for example, one hypothesis of many that could be tested is that AAAG was starting from a lower base of multi-authored papers while DH has included multi-authored papers since its inception

due to its nature), it naturally leads to the question of why the issue of multi-authorship is seen as such a defining feature of DH and why it has featured so prominently in policy documents and advice on tenure and promotion. Without undertaking a much more extensive study, this is difficult to answer but it would be interesting to see, in due course, whether it is the case that DH is distinguished not so much by the fact of multi-authorship but by the nature and rate of the trend towards it? Pertinent aggregate figures for trends across the academic disciplines are included in Wuchty *et al.* (2007 cited above) and note:

Unlike the other areas of research, single authors still produce over 90% of the papers in the arts and humanities. Nevertheless, there is a positive trend toward teams in the arts and humanities ($P < .001$) (p.1037).

Looking to our figures we can see that in CHum single authors produce 69% of papers and in LLC 65%, though a much more extensive study would need to be undertaken in order to systematically explore the issue of whether the trend towards multi-authorships is more statistically significant in DH than in other Humanities disciplines, or whether it is closer to patterns in the Sciences or Arts and Humanities. This would also be interesting in terms of informing other kinds of statistical analysis of DH. For example, Leydesdorff and Akdag Salah (2010) found in terms of a citation impact analysis of DH in ISI's web of science using documents downloaded on 8 September 2009 that

The figure shows that these [DH] documents are cited in a limited domain of two or three groups of journals, namely, new specialist journals with a focus on computer usage in the humanities, and a group of library and information science journals addressing the digitalization of archives and libraries. . . .

The connectivity analysis showed that co-publishing within the discipline (self-defined as 'those who publish in the journal') appears to be more strongly observed in LLC than in CHum. The author connectivity scores show that all three journals display similar characteristics: there is 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. Indeed, this is another perspective on the predominance of single-authorship in DH, as we have noted that a large number of authors published single-authored papers only. Notwithstanding this the connectivity analysis might be a stronger argument for the collaborative nature of DH: while the majority of people do not co-publish, DH has a stronger set of active co-publishers. Another perspective to consider is the 'small discipline' effect (certainly when DH is compared with the discipline of geography): DHers co-publish with a higher proportion of other DHers because it is a small discipline and people are more likely to know each other. The counter argument is that geography is actually a collection of small disciplines with internal co-authorship circles. To pursue these perspectives further, more research is required.

6 Limitations

We mentioned above the necessity of editing the look-up table by hand and this will certainly have created errors. For example, except where authors are personally known to us we are unlikely to have picked up a name change upon, for example, marriage or divorce. However, for our purposes it is unlikely that such errors will have skewed the data to a considerable degree. Such issues could have been tackled with researcher ids but these are not, to the best of our knowledge, applied retrospectively and we did not have the time or resources to assign them based on known and public ids.

We have chosen one journal as a control even though a meta-analysis or combined humanities/social science data set would have presented an overview of publishing trends across the humanities and social sciences. However, the control we have chosen is the most appropriate one at this stage in our research because the DH sample included here covers two journals only and not all available DH publications. So, a single journal is an appropriate control, as it ensures that, as far as possible, we are

comparing like with like. This would not be the case were we to use a combined data set as a control.

7 Conclusion

Rather than comparing percentage shares of papers authored by 1, 2, or more contributors, we have here given absolute observations of how many papers were published. This is appropriate because we are concerned with changes in authorship in the context of an emerging field, where the total number of papers published is also of interest. Whilst the number (and share) of 2+ authored papers have increased over time, it remains true that single-authored papers remain dominant. We feel that this is more apparent using the absolute values analysis.

In this article we have demonstrated that in two core Digital Humanities journals, CHum and LLC, single-authored papers predominate. In CHum we observed a significant increase in dual- and triple-authored papers but not in four- and five-authored papers. In LLC, a significant increase in triple-authored papers can be observed but this is not the case for joint-authored or four- or five-authored papers. In AAAG, single-authored papers are also predominant. In regard to multi-authored papers the statistically significant increases are more wide-ranging than in either LLC or CHum, with increases in all forms of multi-authorship. The connectivity analysis showed that co-publishing within the discipline appears to be more strongly observed in LLC than in CHum. The author connectivity scores show that in CHum, LLC, and AAAG there is a relatively small cohort 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.

Further to these conclusions we have made available some data tables which summarize the data with which we have been working for this article. The tables show author, number of papers, and publication years for each journal. The publication years are given an asterisk where a paper was co-authored by two or more people. Clicking on the header for a column will cause the data to be sorted and redisplayed on the basis of the column contents.

You can find them here: <http://www.ucl.ac.uk/dis/people/oliverdukewilliams/research/hidden-histories>.

Do our findings give an insight into collaboration in DH, as was muted in the tweets that were sent after this article was originally given at DH 2013? For example,

Stats shown by @juliannenyhan on publication patterns. DH maybe not as open and collaborative as self-perceived? #DH2013

— Isabel Galina (@igalina) 19 July 2013

This is difficult to say because of the caveats outlined above in relation to collaboration and multi-authorship. Nevertheless, the methodology and data described here can be seen as one approach of many that might be taken in order to answer such a complex question. Indeed, our findings point to the value of undertaking a wide-ranging qualitative as well as quantitative study to understand concepts, conventions, and manifestations of collaboration and multi-authorship in DH, not only in order to understand more about such practices within DH itself but also to consider further its practices comparatively with other fields.

The research described here can be extended in a number of ways. At the present time we are at work on a study of gender and multi-authorship in DH. We also hope to extend our analysis to include other Digital Humanities journals that were not possible to include here owing to time and resource constraints (for example, *Digital Humanities Quarterly*, *Computing in the Humanities Working Papers* and *Text Technology*, etc), to identify and analyse publications in non-specialist Digital Humanities journals and contributions to, for example, book collections or (as suggested by Geoffrey Rockwell (2013)) to look at conference presentations and their subsequent published forms. Once this has been completed, and a fuller snapshot of digital humanities publishing activity has been taken, we could extend our control to include a meta-analysis or combined humanities/social science bibliographical data set. It would also be interesting to look at some of the data we have gathered in greater detail, for example, does the geographical location of authors seem to impact rates of co-Authorship (due, for example, to

national issues around the evaluation and funding of research?).

As has been argued above joint authorship is but one indicator of collaboration (and, likewise, the fact that a paper has multiple authors is not necessarily indicative of collaboration). In the context of Digital Humanities, it is clear that in order to measure and diagnose authorship, other factors need to be considered in addition to publication patterns, for example, those named on the contributor pages of individual projects.

Clearly there is scope for further quantitative and qualitative research in this area so as to better understand authorship patterns in Digital Humanities.

References

- AAG (n.d.), 'Overview AAG'. <http://www.aag.org/cs/publications/annals> (accessed 1 December 2013).
- Bell, M., Blake, M., Boyle, P. *et al.* (2002). Cross-national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society: Series A*, **165**: 435–64.
- Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, **3**: 4. <http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html/000077.html>.
- Bošnjak, L. and Marušić, A. (2012). Prescribed practices of authorship: Review of codes of ethics from professional bodies and journal guidelines across disciplines. *Scientometrics*, **93**(3): 751–63.
- Bulger, M. E., Meyer, E. T., Flor, G. D. L. *et al.* (2011). *Reinventing Research? Information Practices in the Humanities*. London, UK: Research Information Network.
- Cronin, B. (2005). *The Hand of Science: Academic Writing and its Rewards*. Lanham, MD: Scarecrow Press.
- Deegan, M. and McCarty, W. (2012). *Collaborative Research in the Digital Humanities*. Farnham: Ashgate.
- Engels, T. C. E., Ossenblook, T. L. B., and Spruyt, E. H. J. (2012). Changing publication patterns in the Social Sciences and Humanities, 2000–2009. *Scientometrics*, **93**(2): 373–90.
- Fitzpatrick, K. (2011). *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York: University Press.
- Galina, I. (2013). Stats shown by @juliannenyhan on publication patterns. DH maybe not as open and collaborative as self-perceived? #DH2013. July 19, 2013. Tweet.
- Katz, J. S. and Martin, B. R. (1997). 'What is research collaboration?'. *Research Policy*, **26**: 1–18.
- Kelleher, M. and Hoogland, E. (2011). *Changing Publication Cultures in the Humanities*. Strasbourg: ESF. <http://www.esf.org/research-areas/humanities/strategic-activities/humanities-spring.html#sthash.Ov4wWiI5.dpuf>.
- Koh, A. (2012). 'The Challenges of Digital Scholarship'. *The Chronicle of Higher Education Blogs: ProfHacker*. <http://chronicle.com/blogs/profhacker/the-challenges-of-digital-scholarship/38103>.
- Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980–2000. *Scientometrics*, **58**(1): 35–48.
- Lariviere, V., Gingras, Y., and Archambault, E. (2006). Collaboration networks: A comparative analysis of the natural science, social sciences and the humanities. *Scientometrics*, **68**(3): 519–33.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, **11**(1): 3–15.
- Leydesdorff, L. and Akdag Salah, 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. *Journal of the American Society for Information Science and Technology*, **61**(4): 787–801.
- McCarty, W. (2012). 'Collaborative Research in the Digital Humanities'. In Marilyn, D. and McCarty, W. (eds), *Collaborative Research in the Digital*. Farnham: Ashgate Publishing, Ltd. <http://www.mccarty.org.uk/essays/McCarty,%20Into%20to%20Collaborative%20research%2020111204.pdf>.
- McGrath, L. (ed.) (2011). *Collaborative Approaches to the Digital in English Studies*. Logan, UT: Computers and Composition Digital Press/Utah State University Press. <http://ccdigitalpress.org/cad/CollaborativeApproaches.pdf> (accessed 8 November 2013).
- MLA (2012). *Guidelines for Evaluating Work in Digital Humanities and Digital Media*. http://www.mla.org/guidelines_evaluation_digital.
- Moulin, C. *et al.* (2011). *Research Infrastructures in the Humanities*. Strasbourg: European Science Foundation. <http://www.esf.org/research-areas/humanities/strategic-activities/research-infrastructures-in-the-humanities.html> (accessed 7 July 2012).

- Ramsden, P.** (1994). Describing and explaining research productivity. *Research in Higher Education*, **28**(2): 207–26.
- Rockwell, G.** Digital Humanities 2013. [philosophi.ca. http://philosophi.ca/pmwiki.php/Main/DigitalHumanities2013](http://philosophi.ca/pmwiki.php/Main/DigitalHumanities2013) (accessed 1 December 2013).
- Ruecker, S. and Radzikowska, M.** (2008). The Iterative Design of a Project Charter for Interdisciplinary Research. DIS'08. *Proceedings of the 7th ACM Conference on Designing Interactive Systems*. New York, NY, pp. 288–94. <http://doi.acm.org/10.1145/1394445.1394476> (accessed 1 December 2013).
- Siemens, L., Siemens, R., Cunningham, R. et al.** (2009). INKE Administrative Structure, Omnibus Document. *New Knowledge Environments*, **1**(1): <http://journals.uvic.ca/index.php/INKE/article/view/546> (accessed 1 December 2013).
- Smith, M.** (1958). The trend toward multiple authorship in psychology. *American Psychologist*, **13**: 596–99.
- Spiro, L.** (2009). Collaborative authorship in the humanities. *Digital Scholarship in the Humanities: Exploring the Digital Humanities*. <http://digitalscholarship.wordpress.com/2009/04/21/collaborative-authorship-in-the-humanities/>.
- Spiro, L.** (2011). 'Computing and communicating knowledge: Collaborative approaches to digital humanities projects'. In McGrath, L. (ed.), *Collaborative Approaches to the Digital in English Studies*. Logan, UT: Computers and Composition Digital Press/Utah State University Press. <http://ccdigitalpress.org/cad/CollaborativeApproaches.pdf> (accessed 8 November 2013).
- Stone, S.** (1982). Humanities scholars: Information needs and uses. *Journal of Documentation*, **38**(4): 292–313.
- Subramanyam, K.** (1983). Bibliometric studies of research collaboration: A review. *Journal of Information Science*, **6**: 35.
- Wanner, R. A., Lewis, L. S., and Gregorio, D. I.** (1981). Research productivity in academia: A comparative study of the sciences, social sciences and humanities. *Sociology of Education*, **54**: 238–53.
- Wuchty, S., Jones, B. F., and Uzzi, B.** (2007). The increasing dominance of teams in production of knowledge. *Science*, **316**(5827): 1036–39.