

Comparing Disciplinary Patterns: Exploring the Humanities through the Lens of Scholarly Communication

Daniel Burckhardt <burckhardtd_at_geschichte_dot_hu-berlin_dot_de>, Humboldt-Universität zu Berlin

Abstract

For the past fifteen year, scholarly communication networks such as H-Soz-Kult – the German Information Service for Historians – and H-ArtHist – a specialized discussion and information network for art history based in Germany with an international reach – have been steadily publishing conference announcements and reports. Since both services were born digitally, starting with the listserv infrastructure of the Michigan based H-Net and later supplemented by database-driven web sites, the archives are easily accessible by electronic means. The aim of this paper is to demonstrate that the archives of scholarly communication provide a suitable basis for conducting an assessment of broad fields such as *German historians* or *German art history*, with relatively low technical effort. For the initial analysis of H-Soz-Kult^[1], editorial practices facilitated the automated extraction of the speakers' names as a key feature. But even in cases where no such special markup has been applied, freely available Web services such as AlchemyAPI^[2] provide methods that can be used to achieve comparable results.^[3]

Methods

The starting point for the following analysis was Lev Manovich's provocative talk "How to do Digital Humanities Right?" presented at the Digital Humanities revisited conference in Hannover in December 2013. His initial proposition stated: "don't start with a research question."^[1] Though Manovich's program of cultural analytics makes strong use of visual explorations and focuses mainly on images, the articles presented in the newly founded *Journal of Cultural Analytics* illustrate numerous applications of his concepts to text corpora ranging from literature to restaurant menus [Manovich 2016].

1

By exploring an extensive corpus of texts of a single genre – H-Soz-Kult has published over 5,000 conference reports since 1998 – we were confident of revealing relevant insights into the disciplinary culture of historians as practiced in German-speaking countries over the past fifteen years.

2

The initial approach was primarily technology-driven. Readily available tools for Natural Language Processing (NLP) such as Carrot2 for dynamic text clustering^[2], MALLET for topic modeling [McCallum 2002] and Stanford NER for named entity extraction [Finkel 2005] were applied to the corpus. But the initial outcome was rather inconclusive.

3

As the following example screen shows, Lingo[Osinski 2005] – Carrot2's clustering algorithm – is sensitive to slight term variations such as the German singular and plural of "Edition". Other clusters were labeled with stop words ("Seit," "Denen," "Verschiedene") that carry little information about the actual content within this corresponding set of documents.

4

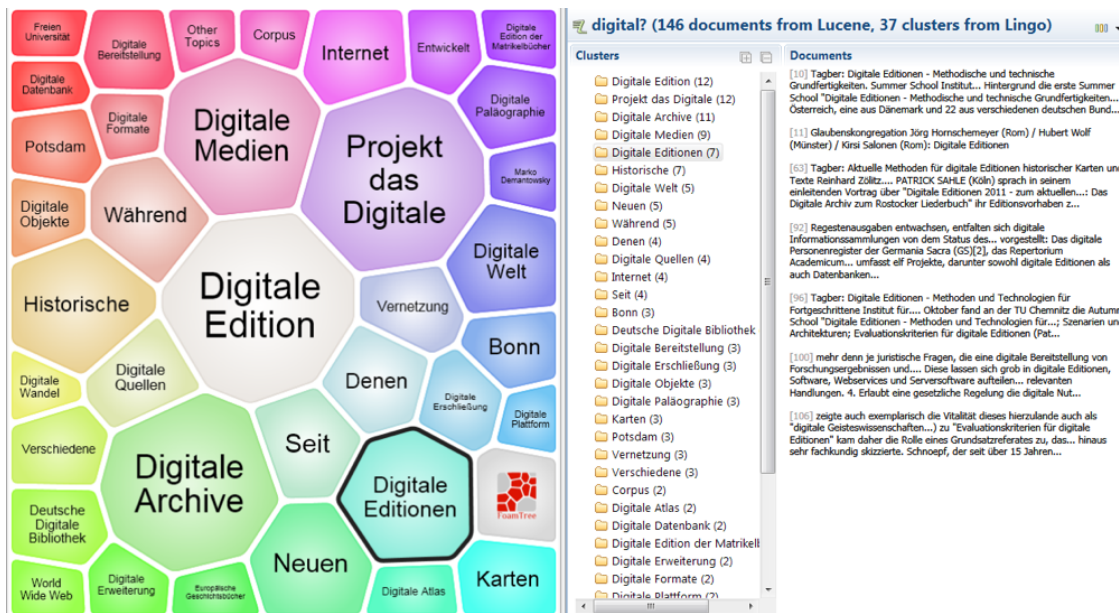


Figure 1. Carrot2's dynamic clustering of Conference Reports matching the query "digital?"

Similar effects could be observed in the list of representative terms for a set of 100 topics generated from the conference reports by an initial run of MALLET: While some of them are immediately recognizable (*"juden jüdischen jüdische jüdischer antisemitismus israel jewish geschichte jerusalem ghetto"*), others again looked like fairly random aggregates of stop words (*"so sondern diese etwa nur dabei gerade damit hat deren"*).

These initial difficulties shouldn't be taken as a general argument against the two clustering methods: As a recent study of scholarly blogging platforms has shown,[Puschmann 2015] it should be possible to improve the results by using optimized stop word lists both for most and least frequent terms in the corpus. In addition, word stemming or morphological analysis and moving to a bilingual topic model (of the 3,537 conference reports since 2008, slightly less than 3,000 are in German, the rest are in English) might avoid enough of the irritations just described and thus make the clusters readable [Mimno 2012].

Stanford NER provides a classifier explicitly trained on large German language sets [Faruqui 2010]. When applied to our corpus, both precision and recall for person names were far from sufficient. Thanks to a lucky editorial decision, it was still possible to get the speakers name extracted from the H-Soz-Kult conference reports by much simpler technical means: From 2008 on, the names of all conference speakers have been emphasized in all upper-case letters (e.g. HANS BLOMME). By using a relatively simple regular expression matching at least two consecutive words in all caps, an extensive list of more than 30,000 speakers lecturing at over 3,500 conferences could be compiled.

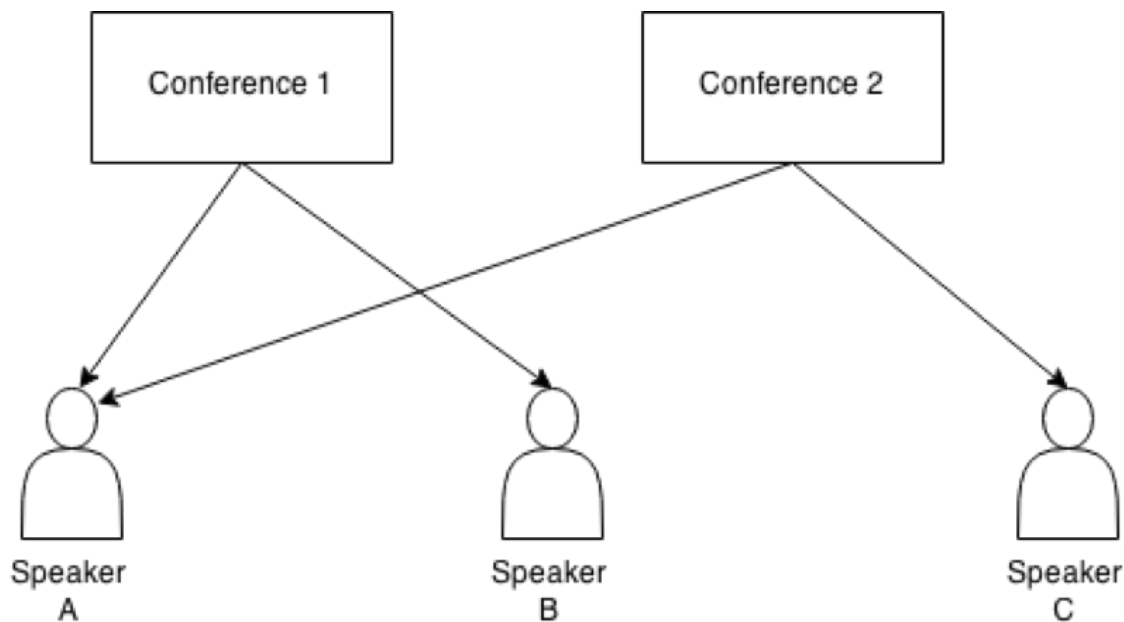


Figure 2. Bipartite network between speakers and conferences. The one-mode projection to speaker nodes leads to (undirected) edges between Speaker A and Speaker B (by common appearance at Conference 1) as well as Speaker A and Speaker C (through Conference 2).

This list of persons grouped by conferences then formed the basis for a bipartite network where each appearance as a speaker defines an edge between a conference and a person. The one-mode projection to the speaker nodes reveals “co-speaker” relations which are structurally similar to co-authorship analysis for scientific papers.

Web-services such as genderize.io^[3] trying to determine the gender of a person based on his or her given name were initially integrated into the extraction pipeline to improve the precision of our regular expression based extraction algorithm: Matches without a first name recognized by this service were flagged for manual review and thus reduced the number of mistakenly extracted person entities. After looking at the gender distribution as one moves from less to more active speakers in our set, we gained valuable insights into the still existing “glass ceilings” in the humanities.

Visualizing the speakers’ network in Gephi^[4] showed some remarkable disciplinary patterns in the field of history, as for example the close-knit groups of early modern and medieval history opposed to the rather frayed borders of contemporary or cultural history.

The Datasets

As mentioned above, the initial dataset consisted of 3,537 conference reports as listed on the website of H-Soz-Kult from January 2008 until summer 2014.^[5] On average, we find slightly more than 40 reports per month with a strong yearly cycle for the conference dates: There is a low due to summer vacation in August, followed by a three months peak conference season after the summer break and a lower but slightly wider peak in spring. These peaks coincide with the preference of putting conferences into the semester breaks of the German academic calendar, excluding the month of August which overlaps with school holidays.

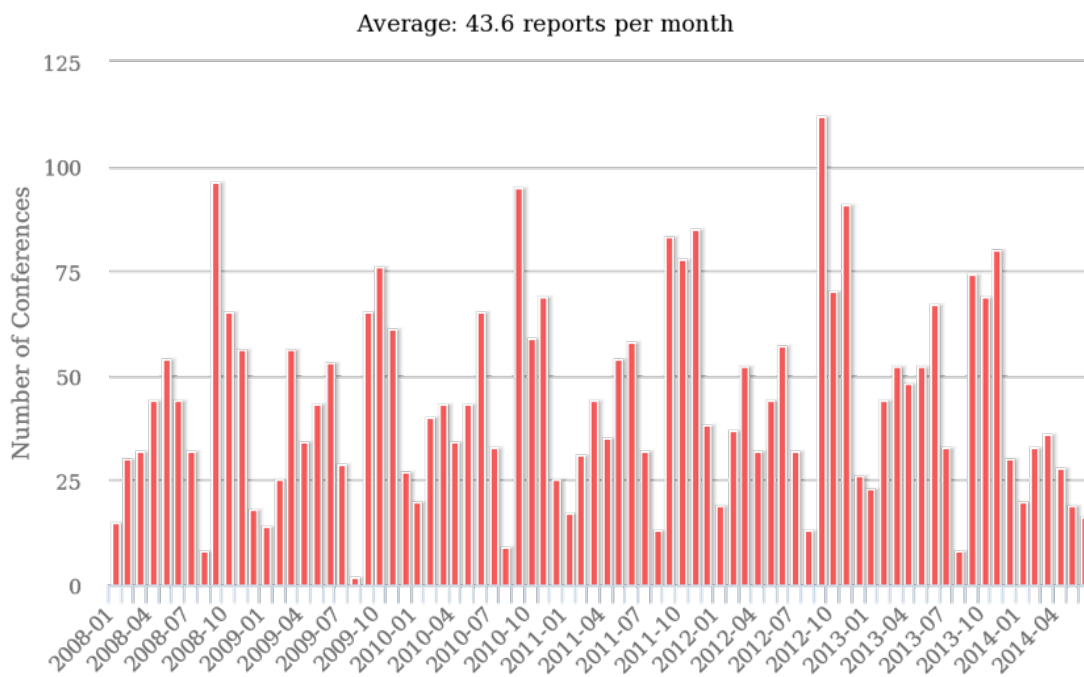


Figure 3. Number of Conferences reported on H-Soz-Kult by Month (January 2008 until June 2014)

The website archive for the second dataset, H-ArtHist's conference announcements, starts a couple years later towards the end of 2010. The monthly average of roughly 39 conferences per month is slightly lower compared to the number of reports on H-Soz-Kult but in a similar range.

12

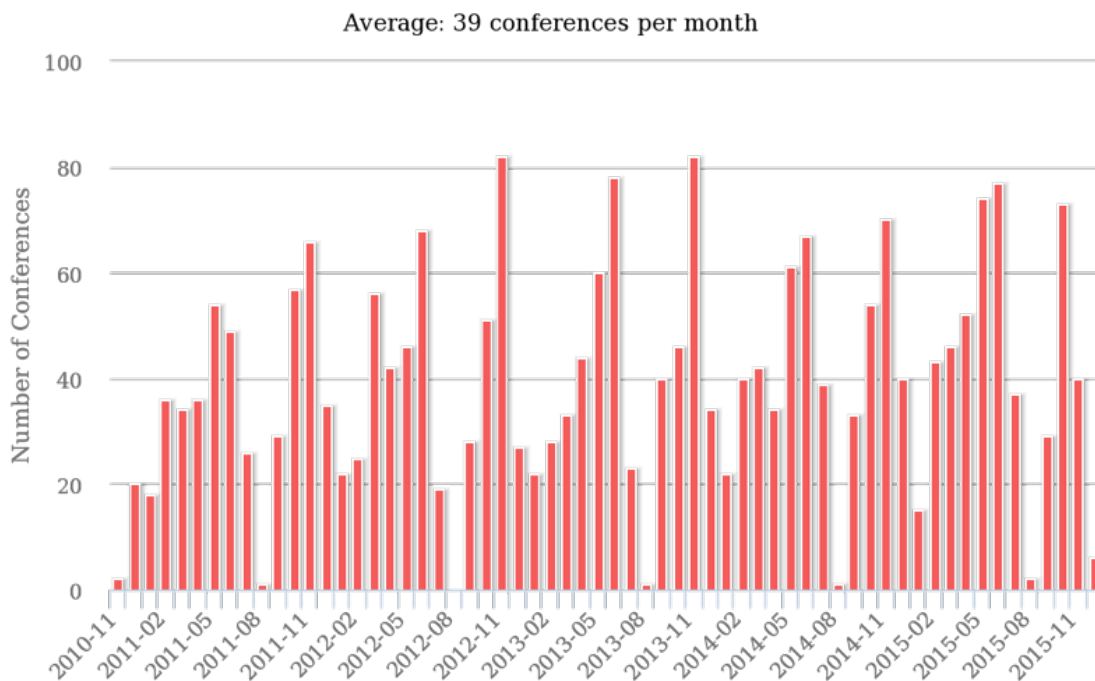


Figure 4. Number of Conferences announced on H-ArtHist by Month (January 2011-December 2015)

The yearly conference cycle shows distinct peaks in November and June of every year and doesn't follow the German academic calendar as closely. This confirms the more international focus of H-ArtHist that we observe among both the countries of subscribers and conference locations:

13

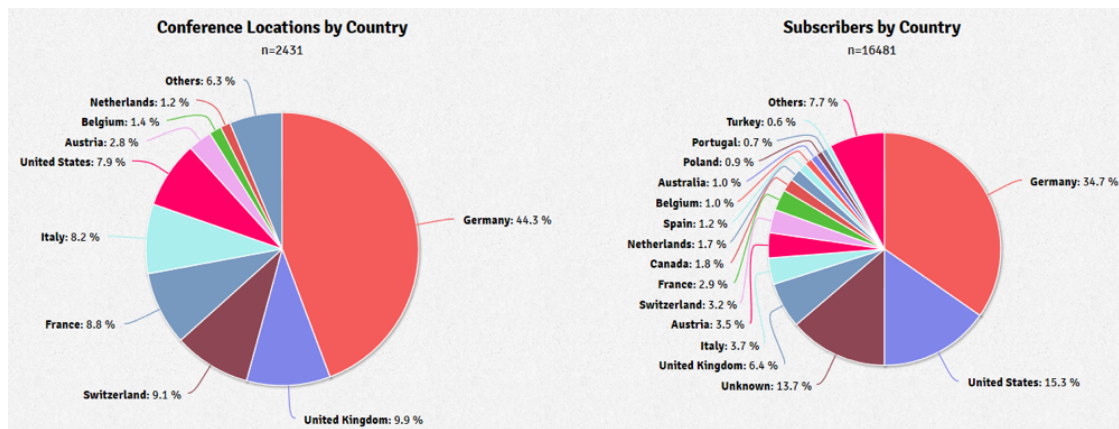


Figure 5. Conference Announcements and Subscribers by Country on H-Arthist

Germany dominates, but we observe a significant share of subscribers and conferences held throughout Western Europe, especially the UK. For the US, we note a striking difference: While more than fifteen percent of the subscribers are based in that country, only 8% of the conferences announced are held in the US. Two factors may account for this difference: many of the influential art history departments belong to private universities, and there is simply no tradition of widely circulating information about conferences. In addition, conferences as announced on H-Arthist are closely tied to a third-party driven mode of research less common in the humanities in North America. For France and Italy, we observe the opposite relation: They show more prominently among conference locations than would be expected from the subscribers' share. This is mainly due to the three independent German art history institutes abroad with an active and actively announced conference schedule, the Kunsthistorisches Institut in Florence, Bibliotheca Hertziana in Rome, and the Deutsches Forum für Kunstgeschichte in Paris.

Preparing the Dataset: Extracting Person Entities

The essential step for building a bipartite network between the conference reports or announcements and the respective speakers was the extraction of the person entities from the unstructured texts. In the first case, the H-Soz-Kult conference reports, a suitable regular expression led to a reasonable set of candidates.^[6]

A significant number of near duplicates resulted from variations of upper and lower case, umlauts, accents and hyphens, illustrated by a single person in our corpus alternatively spelled as Xosé Manoel Núñez, Xosé-Manoel Núñez and Xose-Manoel Nunez. Our solution was to build a "name slug" in lower caps without any accents and just one form of hyphen for any whitespace or hyphen. In the example above, all three cases then resolve to the single key *xose-manoel-nunez*. Another source of slight variations, the trailing S in German genitive (e.g. *ALEXANDER BADENOCHS (Utrecht) Vortrag*), was addressed by the editorial decision to set the trailing s to lower case if it isn't part of the proper noun (*ALEXANDER BADENOCHs (Utrecht) Vortrag*).

We still found numerous misspelled names in the original reports: Between one and two percent of the names contained errors besides variations in accents and punctuation. Since the median of speakers per report is – as will be detailed below – a bit above 10, more than 400 reports had to be manually adjusted. Some of the misspellings could easily be found by checking against a list of known first names (e.g. Ulrike instead of Ulrike). Harder to catch were variations among given names with multiple correct spellings such as Matthias and Mathias. Even for a human, it was sometimes tricky to decide if such variants (e.g. Höppner, Annika vs. Höppner, Anika) needed to be adjusted towards a single person or should remain as is since these are actually two different individuals both correctly spelled.

Both our regular expression as well as more advanced NER algorithms come up with false positives. One could be tempted to start building a classifier that would be trained to distinguish human names from company or room names like BMW AG or "Salle Vasari." This turns out to be very hard for certain edge cases, such as LINDE AG (the company) compared to LINDE NG (a person that actually has a Facebook profile).^[7] A more realistic example from our dataset shows "Brage Bei Der Wieden" as a person's actual name, while "Berge An Der Via Regia" – "mountains along the Via Regia" – is the title of a talk appearing in the conference overview appended to the report. Instead of investing additional technical effort into a programmatic solution, a cursory overlook of the full list of all person entities with manual removal of false positives was used to take care of these cases.

Even after this manual correction step, there are known mistakes in the full list of speakers. Since the effort to manually research the following systematic errors was deemed too high, no attempts were made to unify the following variations:

19

- Changing names over lifetime (Jeannette Madarasz vs. Jeannette Madarasz-Lebenhagen)
- different transcriptions of Cyrillic names into German / English (ow vs. ov)
- inconsistent usage of initials / multiple names
 - Patel, Kiran Klaus vs. Patel, Kiran
 - Miller, Michael vs. Miller, Michael B.

In addition, due to changing affiliations and broad fields of interest that pose a serious obstacle to a systematic disambiguation even by human investigators, we did not try to resolve persons with the same exact name. So ANDREAS SCHNEIDER (Berlin) and ANDREAS SCHNEIDER (Meiningen) as well as HARALD MÜLLER (medieval history) vs. HARALD MÜLLER (law) show up as a single person entity in our dataset. From our experience with the H-ArtHist speakers' list where GND-identifiers were manually researched for the 1,000 speakers with most conference appearances, we can be reasonably sure that namesakes are a relatively rare phenomenon in a corpus with predominantly Western names affecting at most half a percent of the persons.

20

After all these steps, we were left with 30,502 speakers among H-Soz-Kult's 3,537 conference reports. In the case of H-ArtHist, the somewhat messier list resulted in 26,023 persons extracted from 2,459 conference announcements.

21

While the editorial convention in the H-Soz-Kult reports differentiates speakers (e.g. MORITZ FÖLLMER) from persons being spoken about (e.g. Walter Benjamin), this is not the case for H-ArtHist's conference announcements. Therefore, key figures in the historiography of the arts such as Giorgio Vasari, Walter Benjamin and Aby Warburg showed up prominently among the most active persons and have to be manually flagged as non-speaker entities. H-Soz-Kult also differentiates between persons acting as speakers and those welcoming the attendants or taking part in the discussion. Clearly separating these roles can be difficult even for a human reporter: when does a welcome note turn into an introduction; what is the difference of a respondent's statement compared to a far-reaching question during the discussion?

22

In this light, we should expect the number of speakers in conference reports for H-Soz-Kult to be significantly lower than the average number of persons found in H-ArtHist's conference announcements. The actual difference is relatively low (H-Soz-Kult: median of 13, average: 8.62, H-ArtHist: median 14, average: 10.8). This can partially explained by the observation that there are but a limited number of reports on small conferences with just a few speakers (40 reports with 1 to 3 speakers compared to 101 announcements with the same person count).

23

Analyzing the dataset: Estimating the size of a discipline

Sorting the speakers' list by the highest number of conference appearances provided a simple means for identifying the core group of very active (art) historians.

24

It has been shown that preferential attachment – well known conference speakers have a higher chance of being invited to the next conference – can lead to power law distributions. Both for H-Soz-Kult and H-ArtHist, we can observe such a pattern when graphing the distribution of the number of speakers in our list by the number of conference appearances on a double logarithmic scale:

25

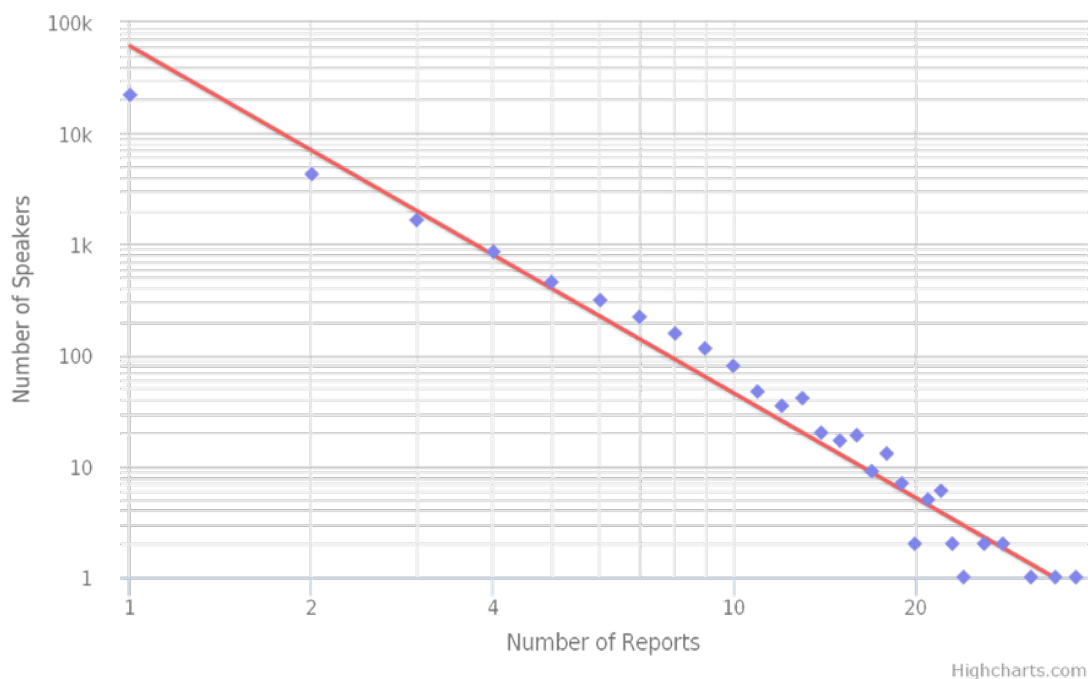


Figure 6. Speakers' Distribution by Number of Appearances on H-Soz-Kult (Double Logarithmic)

In the case of H-Soz-Kult, there are more than 22,000 persons appearing in just a single report. Around 4,300 appear in two reports, 1,650 in three reports. On the very opposite, we will find but 23 people appearing twenty times or more. To the very right of this distribution, we find a single person who presented at 37 conferences within less than seven years. Very prominent among this highly active group are the leaders of non-university research institutes. If we focus on the group of the most active university professors, we see that younger researchers clearly dominate over eminent scholars towards the end of their academic career.

Twenty-two thousand persons appearing at just one conference over more than six years seems to be an incredibly high figure. Picking out random samples, it turns out to be a very diverse set ranging from graduate students presenting their topic for the first time, visiting scholars from abroad, scholars from neighboring disciplines as well as politicians and other well known but rather rare speakers in an academic setting. More important: The total number of 30,000 people is in a similar range as the total number of subscriptions to H-Soz-Kult which is currently around 25,000 people. Due to the wide variety of professional backgrounds just mentioned, we wouldn't expect a full overlap. But we can safely conjecture that of the roughly 8,000 people appearing twice or more, a significant amount will also turn out to be subscribers to the mailing list. This group forms a "core" of research oriented historians in the German speaking countries.

DE-Statistik, the Federal Statistical Office, counts slightly more than 6,000 full- or part-time employees at German Universities in 2013; accounting for a comparable per capita share in Austria and the German speaking part of Switzerland, we'd expect a bit more than 7,000 full- or part-time employees in history in the German speaking academy [DeStatis 2014].

For H-ArtHist, the number of entities extracted from the announcements amounts to slightly more than 26,000 persons with roughly 21,000 appearing just once. Due to the less accurate name extraction process and no manual filtering so far, we should take this number with a pinch of salt. But if we again assume the remaining 5,000 people that appear more than once to represent the academic core of art history; assume that half of them are active in the German speaking academy (based on the country distribution of the announcements), we can estimate that there are about 2,500 academic art historians, a count again not too far from the employment figures reported by DE-Statistik.^[8]

Analyzing the dataset: The Gender of the Discipline

As mentioned above, checking the first word of extracted person entities against a given set of well known given names was initially introduced to identify typos and false positives. Since genderize.io provided a suitable API for this job, we ended up with an automated assignment of female or male gender for the wide majority of persons extracted from our datasets.

Gender inequality in the academia has been extensively studied for the past fifty years.^[9] Lyndal Roper, in her keynote at the 2014 German Historikertag in Göttingen stated that 45% of academic staff in history departments in Germany is female.^[10] When one looks at the number of professors, this number drops to 29%.^[11]

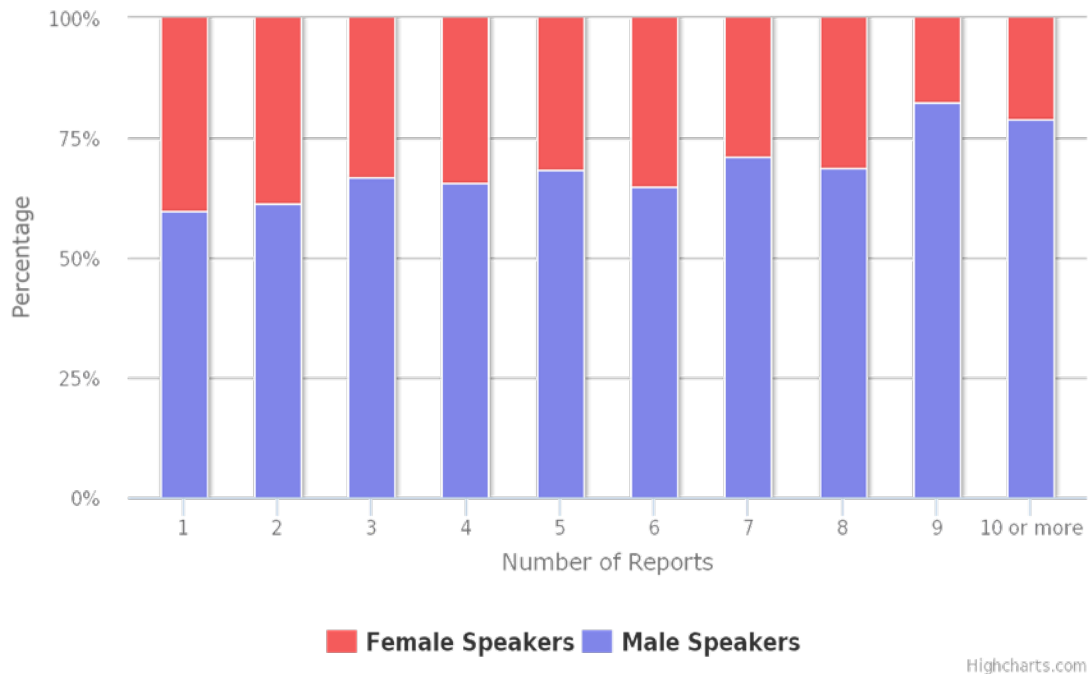


Figure 7. Relation between Female and Male Speakers by Number of Conference Reports on H-Soz-Kult

If we look at the figures from our reports, we find a “Leaky Pipe” between occasional and regular conference speakers similar to the one between University employment and professorship. Overall, slightly less than 40% of the speakers are female, not too far below the 45% figure of female academic staff. If we focus on the more active set of speakers appearing at five or more conferences, the female percentage drops to 27% (309 female vs. 808 male). If we focus solely on the most prominent speakers with 10 or more talks, this number further decreases to 25%. To put this percentage into perspective, note that we find an average of around 50 habilitations in history in Germany per year for the past five years, roughly a third (16.6) by women. The reasons for this enduring gender gap in German history departments have been concisely described by Barbara Stollberg-Rillinger – one of the only two women appearing in our otherwise all-male group of speakers with 20 or more appearances – in a panel discussion celebrating thirty years of the DFG’s Gottfried Wilhelm Leibniz Prize: “The sudden drop [among women] happens in the postdoc phase, when family planning is pending, but the academic system only offers uncertain perspectives. Female academics that have to muddle through temporary and poorly paid third-party projects until the age of 40 are often forced to choose between a child and a career.”^[12]

Art history, maybe not that unexpectedly, shows a significantly smaller gender gap: The distribution among all persons mentioned in our announcements is 56% (male) to 44% (female). If we only take into consideration the people being mentioned in five or more announcements, the woman’s percentage drops slightly to 40%, for 10 or more announcements to 35%. This is roughly comparable to the share of women’s habilitation in German art history.^[13] What our dataset fails to convey is the fact that the relatively even picture looks very distorted once we consider the major loss in the field of art history earlier on the career track: Around 85% of all degrees at the BA and MA level in art history in Germany are completed by female students, and still more than three quarters of all PhDs are successfully defended by women.^[14] But we find significantly fewer women moving from advanced education to advanced research compared to male graduates in the field, both measured by conference appearances as well as by employment figures.^[15]

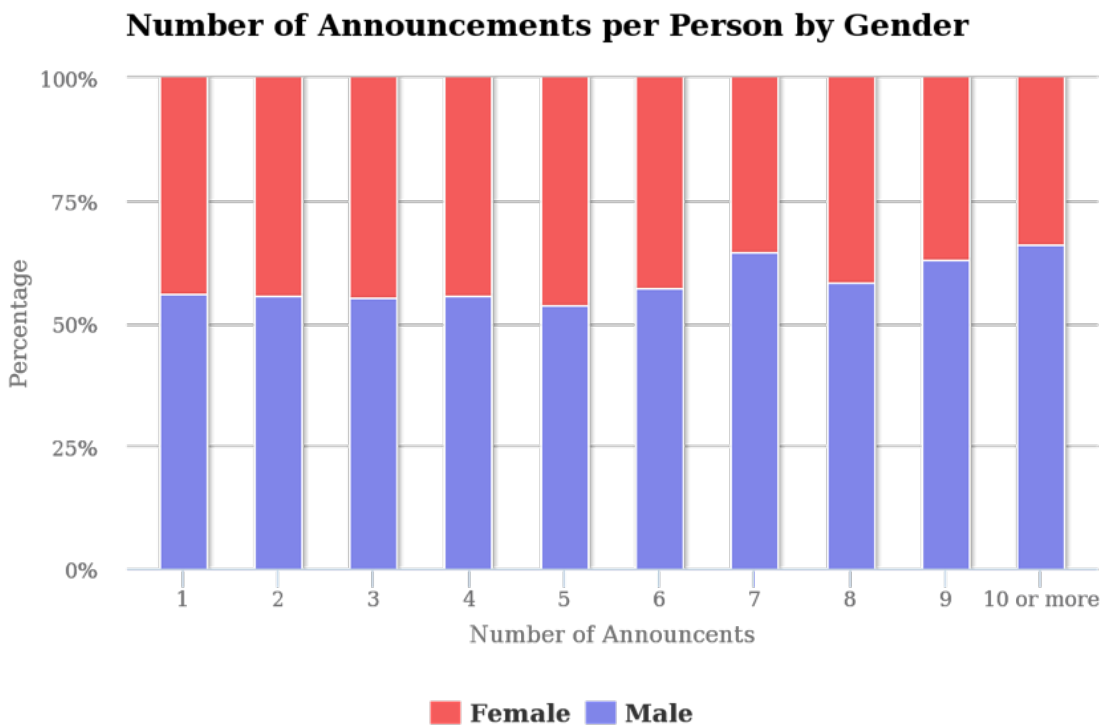


Figure 8. Relation between Female and Male Speakers by Number of Conference Announcements on H-ArtHist

Analyzing the dataset: From Text to Social Networks

After these rather simple but nevertheless insightful counts of basic properties, we finally move over to the more complex relations among persons studied through the network of “co-speaker” relations. Since this network is defined by drawing a tie between any two persons speaking at the same conference, we need to ask if it is reasonable to assume that this tie implies some kind of social relation between the two?^[16] As discussed before, a typical conference consists of a bit more than ten speakers. So while we can never be sure that a social interaction between all of the around 80 possible pairings within such a group actually took place, the assumption, that there is at least a strong interaction potential as part of the discussions, during the breaks in-between sessions, as well as during a possible conference dinner, seems valid. And even in cases where no such interaction took place, we can still assume some sort of shared research interest since most conferences are organized around a unifying theme.

34

If we compare our tie to those commonly studied in bibliometrics, co-speakership is certainly weaker than co-authorship, a relation we still only very rarely observe in the humanities. On the other hand, our tie seems quite a bit stronger than citations among different authors; a primarily intellectual connection among them, which – at least in larger disciplines – only rarely implies a real world social interaction.^[17]

35

Since speakers that only appear once don’t add any additional connections between otherwise unrelated speakers, we started with the reduced network including only persons appearing at two or more conferences. In the case of H-Soz-Kult, we get a graph with 8,361 speaker nodes connected by more than 130,000 edges. For H-ArtHist, the corresponding graph consists of 5,165 person nodes and a bit less than 65,000 edges.

36

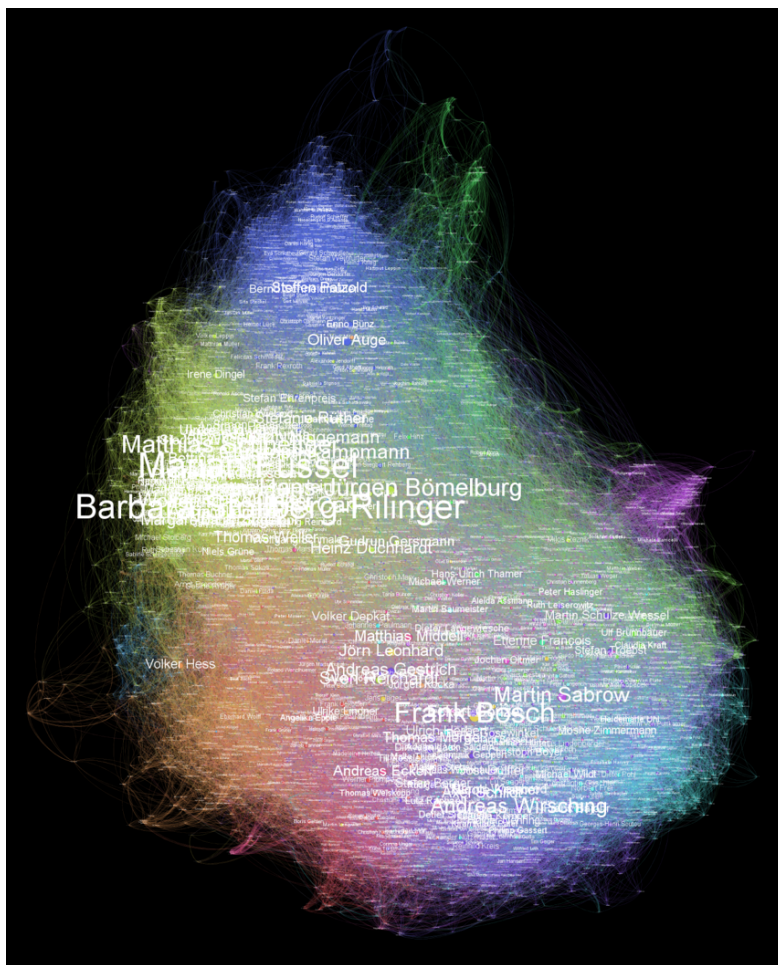


Figure 9. Complete graph for H-Soz-Kult colored by modularity classes, label sizes by Eigenvector centrality

Though the full pictures shows the typical hair-ball one would expect, coloring the graph by modularity classes, Gephi's community detection algorithm^[18], shows a surprising feature for the case of German historians: We find a recognizable tightly knit group – in our case in green on the upper left – which can be easily identified as historians of the early modern period. On top of them (in this graph in blue), we find medieval history. Towards the bottom, we find researchers related to contemporary history, colored in the purplish blue. The multiplicity of historical research still organized primarily along epochs, is the one feature that stands out from a distance in this graph.

37

No such division could be observed in the corresponding Gephi-graph for art history. We find a few well known scholars in the center of the graph, but their neighborhood is much less defined by specific periods or topics than by the geographical location of the research institute.

38

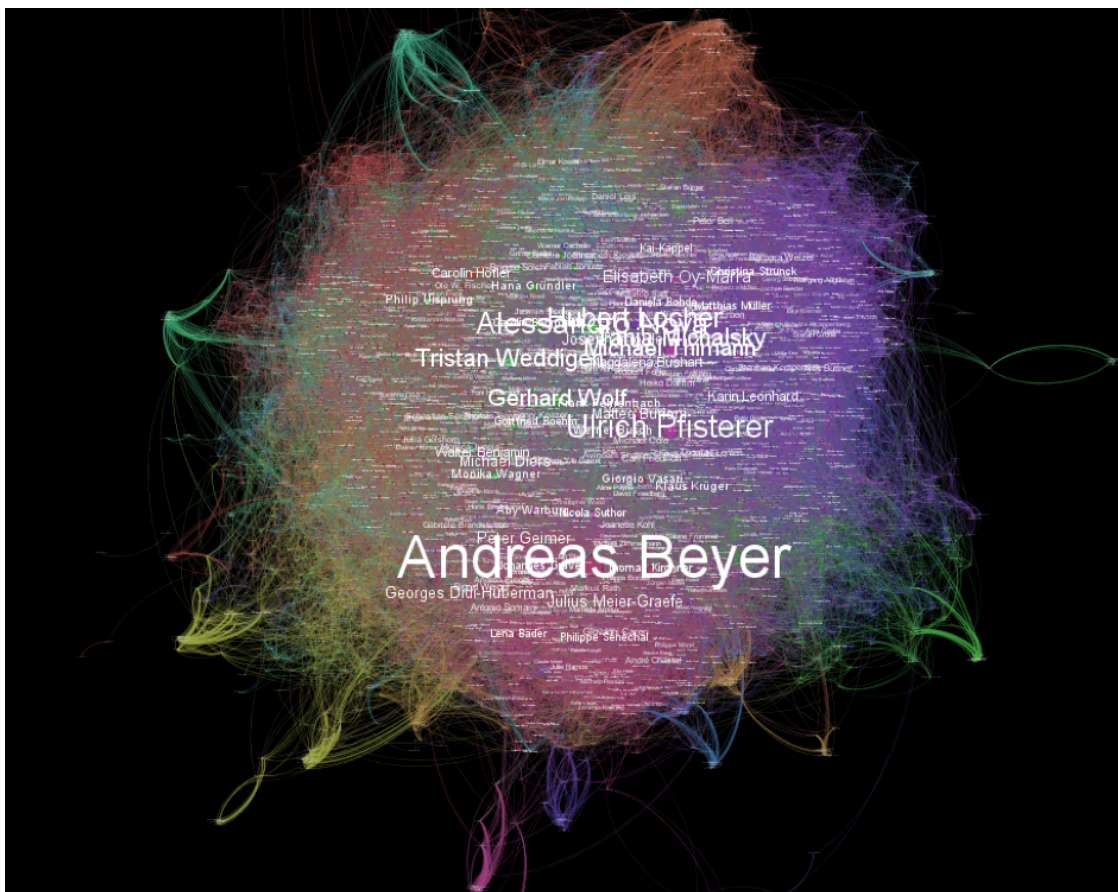


Figure 10. Complete graph for H-Arthist colored by modularity classes, label sizes by Eigenvector centrality

The more unitary structure of the narrower field of art history is probably due to the smaller sized institutes and departments resulting in wider teaching duties and a lower degree of specialization.

39

Of the commonly studied centrality measures for one-mode networks, Degree distribution and Betweenness centrality directly correlate with the total number of conference appearances of a person in our datasets and therefore add no additional insights to our analysis.^[19] The one measure that led to interesting insights is the so called Eigenvector centrality, a measure “assessing how well connected an individual is to the parts of the network with the greatest connectivity.”

40

Going back to the full network of H-Soz-Kult, one sees that the early modern period shows a group of highly connected people (marked by the size of the labels). This corresponds to the gateway function between the medieval and modern periods as well as the tight connections among themselves. So it is absolutely not surprising that the only working group of the German Historian’s association organized around a specific period is the “Arbeitsgemeinschaft Frühe Neuzeit.” The opposite phenomenon can be observed among scholars from the younger field of Transnational History and the recently emerging field of “Visual History.” The Eigenvector centralities of the most prominent actors from these fields are significantly lower than those of comparable speakers from more established fields showing a similar number of conference appearances. While a high centrality core comparable to the early modern period is lacking in the case of art history, we find similar patterns for the group of people with a low centrality despite of frequent appearances: Australian Aboriginal Art or the Materiality of Magic are clearly topics far off the beaten paths in the discipline.

41

Another fact worth investigating is the number of conference-“buddies”. Is it common to have two speakers that have lectured together at one or more conferences before? A look at the graph from the H-Soz-Kult set below shows, that edge weight is a rapidly decreasing distribution: In average, just two speakers at a conference have lectured together before or will do this on a follow-up conference in our set. Only six “couples” share six, and a single pair belonging to the same University department appeared at eight conferences together. Enduring relations seem to be a rare exception rather than the norm.

42

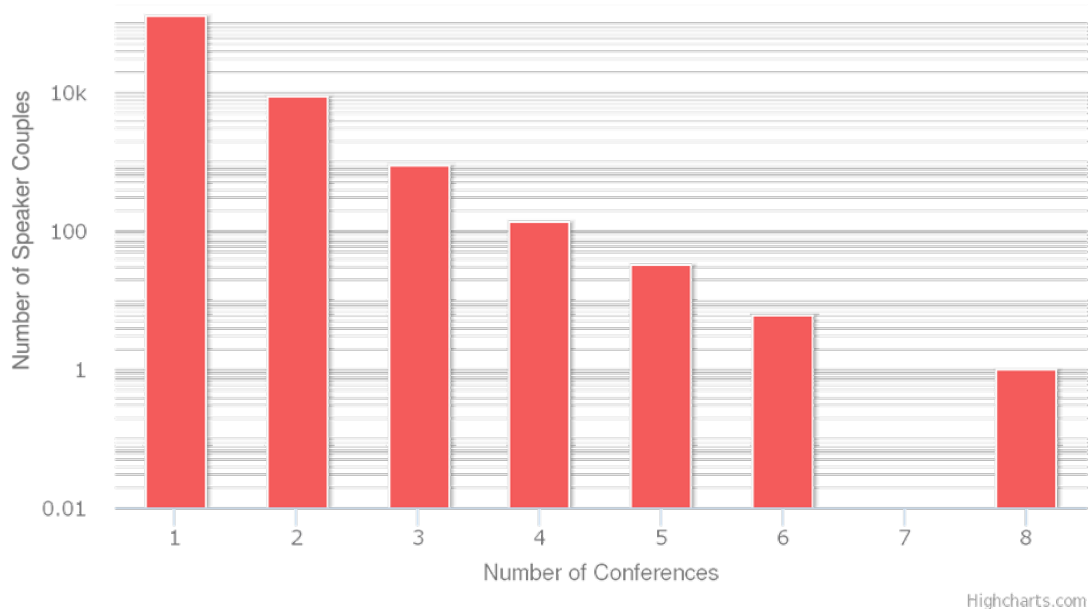


Figure 11. Number of shared Conference Appearances on H-Soz-Kult (Logarithmic Scale)

This relative rareness of co-appearances leads to a very simple but amazingly effective clustering algorithm for “socially similar” conferences: If two conferences share at least two speakers, there is a high chance of thematic overlap between these events. Contrary to content-based similarity measures, this metric is independent of the languages of the report: “Atom global: Kulturen und Krisen im Vergleich” is close both to the English language report on “Accidental Armageddons: The Nuclear Crisis and the Culture of the Second Cold War, 1975-1989” as well as the German report on “In Bewegung. Neue Geschichten der Umweltbewegungen.”

Conclusions

Due to the rise of electronic publications and communications (e-mail, Twitter), large scale social network analysis will become more and more feasible, especially in politics, contemporary history as well as in science studies. As the case of H-ArtHist has shown, readily available NER-services such as AlchemyAPI are working reasonably well to get roughly comparable results on corpora where implicit (H-Soz-Kult) or explicit (well structured TEI-editions) markup of persons is missing.

The data we’ve been analyzing covers the past six (H-Soz-Kult) and five (H-ArtHist) years respectively. In the humanities, this corresponds to a relatively brief period roughly comparable to the time used for writing a PhD or a habilitation thesis. Therefore we cannot yet trace personal careers all the way from the first conference appearance – usually at the beginning of a PhD – all the way up to the appointment as a professor generally more than a decade later. Hopefully, both services continue their work in order to repeat this analysis in a few years with a special focus on newly appointed faculty. Such an analysis might identify successful personal and thematic networks similar to the “Arbeitskreis Historische Frauen- und Geschlechterforschung” recently studied by Angelika Schaser [Schaser 2015]^[20].

While citation indices and impact factors are widely accepted in many branches of science, social sciences and economics, we are well aware of the fact that attempts to quantify and measure institutional or personal research activity have been strongly opposed by professional organizations in the humanities, especially in Germany.^[21] This dislike can at least partially be attributed to the incomplete measures of currently available indices: Frank Bösch, the most prominent German historian in our dataset, is listed with only 13 publications and 12 citations in Thomson Reuter’s Web of Science. The resulting h-index of only 2 would probably be too low to get tenure recommendation in most disciplines that value these measures. Once we perform a similar search through Google Scholar which accounts for monographic citations as well, both the number of publications and the h-index rise by a magnitude to around 150 papers and an h-index of 17. As this single example shows, an adequate corpus is crucial for any serious attempt to quantify academic performance.

But even where a suitable corpus is available, professional curiosity eager to reveal fine grained images by assembling pieces of public data must not forget the right of individuals not to be spied upon.^[22] For this reason, we are still not sure if the two

datasets discussed in this paper should be treated as Open Data, free for anyone interested to investigate or if the preferred mode of access should limit it to researchers agreeing to value the privacy of all persons involved.

Notes

[1] First presented at the Historical Network Research Conference 2014 in Ghent.

[2] AlchemyAPI Entity Extraction, <http://www.alchemyapi.com/products/alchemy/language/entity-extraction>.

[3] Thanks to Victoria H. Scott for suggesting analyzing H-ArtHist's conference announcements in similar ways.

[1] <http://lab.softwarestudies.com/2013/12/how-to-do-digital-humanities-right.html>

[2] Carrot2, open source framework for building search clustering engines, <http://project.carrot2.org/>

[3] Genderize.io, Determine the gender of a first name, <https://genderize.io/>

[4] Gephi: The Open Graph Viz Platform, <https://gephi.org/>

[5] See <http://www.hsozkult.de/conferencereport/page>

[6] `\b(\p{Lu}[\p{Lu}\x{00df}–\']+\s[\p{Lu}\x{00df}–\']\s.)+[\p{Lu}\x{00df}])s?\b`

`\p{Lu}`, short for `\p{Uppercase_Letter}`; a Unicode upper case `\x{00df}`: German ß (has no upper-case variant) optional lower case at the end for genitive (HANS MEIERs Vortrag).

[7] Almost any general assumption about person's name will break down, especially as one leaves the Western world, <http://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/>

[8] For art history, DE-Statistik counts 1,700 full- or part-time employees, so roughly 2,000 in the German speaking countries.

[9] Linda Nochlin's groundbreaking 1971 article "Why have there been not great women artists" contains many answers to the related questions of "Why have there been no great women [art] historians?" For an overview of recent studies focusing on teaching and research in the US see [Savonick 2015]

[10] [Roper 2013]. For an up-to-date history of women in West German history departments, see [Hagemann 2016]

[11] 2,812 of 6,137 (45%) employees and 225 of the 753 (29%) professors in history in Germany are female according to DE-Statistik, Personal an Hochschulen 2013.

[12] Quoted in NN. 01.04.2015. Im Jahr des Mannes. Hat der Leibniz-Preis ein Frauenproblem? *Frankfurter Allgemeine Zeitung*, N4.

[13] 4 out of 12 habilitations in art history in Germany in 2013, or 60 out of 124 if we take the more reliable period between 2004 and 2013, DE-Statistik, Personal an Hochschulen 2013.

[14] 1,261 out of 1,463 Bachelor-degrees, 535 out of 639 MAs and 126 out of 166 PhDs were finished by woman in 2013. See DE-Statistik. 2014. *Prüfungen an Hochschulen 2013*.

https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PruefungenHochschulen2110420137004.pdf?__blob=publicationFile

[15] 1,530 of 2,755 (56%) employees and 167 of the 405 (41%) professors in art history are female according to DE-Statistik, Personal an Hochschulen 2013.

[16] For a much more in depth discussion of these aspects, see [Lemerrier 2012]

[17] For details concerning these two types of networks, see [Havemann 2009].

[18] Gephi: The Open Graph Viz Platform, <http://gephi.github.io/>. The community detection is based on the algorithm outlined in [Blondel 2008]

[19] A thorough introduction to bi- and multimodal networks is given in Scott Weingart's "Networks Demystified Series" part 9: Bimodal networks, <http://www.scottbot.net/HIAL/?p=41158>. He points out that Gephi, our preferred tool for Social Network Analysis, is not really suited to calculate centrality measures for such networks, pointing to Tore Opsahl's detailed overview of Clustering in two-mode Networks, <http://toreopsahl.com/tnet/two-mode-networks/clustering/>

[20] Of the 18 co-founders of this network in 1990 at the PhD level at that time, thirteen are now tenured professors.

[21] See for example [Prinz 2009].

[22] This ethical tension has been investigated by [Berendt 2015]

Works Cited

- Berendt 2015** Berendt, B., Büchler, M., Rockwell, G. "Is it Research or is it Spying? Thinking-Through Ethics in Big Data AI and Other Knowledge Sciences", *KI - Künstliche Intelligenz*, 29.2 (2015): 223–232. <http://link.springer.com/article/10.1007/s13218-015-0355-2> (2015)
- Blondel 2008** Blondel, V., Guillaume, J.-L., Lambiotte, R., Lefebvre E. "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, 10, P1000 (2008)
- DeStatis 2014** DE-Statistik. *Personal an Hochschulen 2013*. https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PersonalHochschulen2110440137004.pdf?__blob=publicationFile (2014)
- Faruqui 2010** Faruqui, M., Pado, S. "Training and Evaluating a German Named Entity Recognizer with Semantic Generalization", *Proceedings of KONVENS 2010*, Saarbrücken, Germany. http://www.nlpado.de/~sebastian/pub/papers/konvens10_faruqui.pdf (2010)
- Finkel 2005** Finkel, J., Klein, D., Manning, C. *Stanford Named Entity Recognizer (NER)*. <http://nlp.stanford.edu/software/CRF-NER.shtml> (2005)
- Hagemann 2016** Hagemann, K. "Gleichberechtigt? Frauen in der bundesdeutschen Geschichtswissenschaft", *Zeithistorische Forschungen/Studies in Contemporary History* 13.1, <http://www.zeithistorische-forschungen.de/1-2016/id=5333> (2016)
- Havemann 2009** Havemann, F. *Einführung in die Bibliometrie*. Berlin: Gesellschaft für Wissenschaftsforschung (2009).
- Lemercier 2012** Lemercier, C. "Formale Methoden der Netzwerkanalyse in den Geschichtswissenschaften: Warum und Wie?", *Österreichische Zeitschrift für Geschichtswissenschaften*, 23/2012/1, 16-41 (2012)
- Manovich 2013** Manovich, L. "How to Do Digital Humanities Right?" *Digital Humanities Revisted* Herrenhausen Palace, Hanover/Germany, December 6, 2013.
- Manovich 2016** Manovich, L. "The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics", *Journal of Cultural Analytics*. <http://culturalanalytics.org/2016/05/the-science-of-culture-social-computing-digital-humanities-and-cultural-analytics/> (2016)
- McCallum 2002** McCallum, A. K. *MALLET: A Machine Learning for Language Toolkit*. <http://mallet.cs.umass.edu> (2002)
- Mimno 2012** Mimno, D. "Computational historiography: Data mining in a century of classics journals", *Journal of Computing and Cultural Heritage*, 5.1.3 (2012). <http://doi.acm.org/10.1145/2160165.2160168>
- Osinski 2005** Osinski, S., Weiss, D. "A concept-driven algorithm for clustering search results", *IEEE Intelligent Systems*, 20.3 (2005): 48-54.
- Prinz 2009** Prinz, C., Hohls, R. "Qualitätsmessung, Evaluation, Forschungsrating: Risiken und Chancen für die Geschichtswissenschaften?", *Historisches Forum*. Berlin: Humboldt.-Universität (2009).
- Puschmann 2015** Puschmann, C., Bastos, M. T. "How Digital are the Digital Humanities? An Analysis of Two Scholarly Blogging Platforms". *PLoS ONE*, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115035> (2015)
- Roper 2013** Roper, L. "Opening Remarks, Winners and Losers", *The 50th German Historical Association Historiktag conference*, September 23 (2014)
- Savonick 2015** Savonick, D., Davidson, C. N. "Gender Bias in Academe: An Annotated Bibliography of Important Recent Studies", *Hastac*, January 26. <https://www.hastac.org/blogs/superadmin/2015/01/26/gender-bias-academe-annotated-bibliography-important-recent-studies> (2015)
- Schaser 2015** Schaser, A. *Der Arbeitskreis Historische Frauen- und Geschlechterforschung 1990 bis 2015. Wissenschaftliche Professionalisierung im Netzwerk*. Hamburg: Druckhaus Köthen (2015)



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.