Technische Universität Berlin

Lockdown Bibliometrics: Papers not submitted to the STI Conference 2020 in Aarhus

Sociology of Science Discussion Papers
SoS Discussion Paper 2/2020

10 September 2020

Lockdown Bibliometrics: Papers not submitted to the STI conference 2020 in Aarhus
Editor: Jochen Gläser
SoS Discussion Paper 2/2020
TU Berlin
Social Studies of Science and Technology
2020

Table of contents

Preface

This discussion paper presents the contributions we would have submitted to the STI Indicators conference, which was planned to be held from September 2nd–4th in Aarhus, Denmark. The conference was postponed until September 2021 due to the Covid-19 pandemic. The decision was made when most of the papers we intended to submit were near completion and all of us had already invested a significant amount of work. Since we all will want to submit new work to next year's conference, we thought it a good idea to make our work available in this discussion paper.

Although publishing bibliometrics papers in a "Sociology of Science" discussion paper series might seem unusual, all four papers deal with sociological foundations of bibliometrics or applications of bibliometric methods in sociological research. Grit Laudel developed and tested bibliometric measures that could be used as indicators of a researcher's inclusion in their international scientific communities, thereby contributing to a historical-sociological project on the changing inclusion of East German researchers in their scientific communities after German unification. Matthias Held et al. used a ground truth of research topics that were described by researchers working on them to test bibliometric approaches to topic reconstruction – with rather depressing results. Jochen Gläser reviewed bibliometricians' approaches to "expert validation" of bibliometric maps from the perspective of sociological methodology and suggested an alternative approach. Finally, in a paper that was originally meant to be a poster, Jenny Oltersdorf et al. presented a workflow for constructing bibliometric data bases from scratch, which they are currently applying inthe construction of a publication and citation database that will enable the study the link between knowledge production and communication in German art history and international relations.

We can present our papers but their discussion is up to readers. We invite readers to send us their feedback and intend to update the discussion paper with the feedback and our responses.

Berlin, 1 September 2020 Jochen Gläser

Can we find Bibliometric Traces of the Inclusion of Researchers in their Scientific Communities?

Grit Laudel

grit.laudel@tu-berlin.de, TU Berlin, FH 9-1, Fraunhoferstr. 33-36, 10587 Berlin (Germany)

Introduction

The original idea of scientific communities developed by Robert Merton, Thomas Kuhn and Michael Polanyi implied the idea that all members of scientific communities equally participate in the production of scientific knowledge (Kuhn 1970, Merton 1973, Polanyi 1962). Meanwhile, an increasing body of research has demonstrated that researchers differ in the degree to which they are included in their international scientific communities. On one end of the spectrum we find researchers who cannot even fully access the state of the art of their community, on the other end we find those who contribute knowledge claims which are used by others and who actively shape the research directions of their community. The unequal inclusion of researchers is a ubiquitous phenomenon. It is often discussed in terms of a North-South divide. However, we find different levels of inclusion among researchers of the Global North, too. Gender biases and unequal access to resources are just two reasons for these differences.

Although the phenomenon is not new, little attention has been paid to the methodological challenges of studying it. Bibliometric studies so far only used few indicators such as publications in non-mainstream and mainstream (SCI-indexed) journals and their interdependence, which was studied for peripheral Asian scientific communities (Davis and Eisemon 1989) and for South Africa (Tijssen 2007). Arunachalam and Manorama (1988) studied publications in non-mainstream journals in India in terms of the age of their references and the cited journals. In a recent more advanced study, Chavarro et al. (2017) discussed the publication strategies of Colombian researchers who publish in non-mainstream journals and the links of these researchers to the mainstream of their scientific community.

Although there is only a limited number of studies, they point to the potential of bibliometric indicators for investigating the inclusion of researchers in their scientific communities. In this paper, I present results of a study that further explores the extent to which bibliometric methods can contribute to characterize researchers' inclusion into their scientific communities. This attempt is part of a larger sociological and historical study on the inclusion of East German scientists before and after German unification. Since political and resource conditions for East German researchers changed drastically after unification, this is an ideal case for studying the long-term dynamics of inclusion.

Theoretical concepts and their operationalization

Inclusion is understood here as the extent to which a researcher participates in their community's knowledge production. Participation in knowledge production requires various interactions between a researcher and their fellow community members, which can be understood as dimensions of inclusion. I draw on our own previous work which distinguishes between:

- access to a community's knowledge claims (through participation in formal communication – publications, conferences – and informal communication),
- submission of research contributions to the community (in publications),
- perception and utilization of these contributions by other community members,
- participation in collaborative knowledge production processes,
- participation in a community's decision-making processes (e.g. as a reviewer of contributions and research proposals, member of editorial boards, or recruitment committees) (Gläser and Laudel 2001[1]).

Not all these dimensions can be measured with bibliometric methods. In this paper, I focus on those dimensions which can at least partially be measured by bibliometric indicators (Table 1). A researcher's *access to published knowledge* can be partially deduced from properties of the references in the researcher's publications. The journals cited, community members cited and age of references provide some indication of barriers to access. Researchers who have no easy access to the published literature may refer to older literature, as previous studies have shown (Arunachalam and Manorama 1988). *Offers of research contributions* can be described by the number of publications, and their visibility is indicated by the visibility of the journal in which they appear to the scientific community. Researchers who publish their findings in highly visible journals have better chances that their publications are noticed, read, and used. The actual *perception of a researcher's contributions* can be determined by the numbers of citations and by identifying community members who cite them. For my study, perceptions of contributions by researchers from the Eastern bloc versus researchers from the Western bloc are of particular interest. Note that I interpret citations conservatively, i.e. as an indicator of a publication being noticed. Finally, researchers can be *included in collaborative research processes*, which becomes bibliometrically visible if this collaboration is rewarded by a co-authorship (see Laudel 2001 on the limitations of this indicator).

[1] In our previous work we used the term "integration" rather than "inclusion". The term "inclusion" can be better connected to the rich sociological theory on inclusion in various spheres of society.

Table 1. *Dimensions of inclusion and their bibliometric operationalization in this study*

Concept	*Operationalization by bibliometric indicators*
Access to published knowledge	Journals cited, community members cited, age of references
Offers of research contributions to the community in publications	Number of publications in SCI journals; visibility of journal in which researchers publish
Perception of contributions by the scientific community	Number of citations, location of researchers citing the publications (here: scientific communities from the Western bloc and the Eastern bloc)
Inclusion in collaborative research processes	Co-authored publications

Approach

To test whether these indicators can reveal patterns of inclusion of East German researchers, I conduct three comparisons, namely

- a comparison of researchers' inclusion before (1980-1990) and after German unification (1991-2000),
- a comparison of levels of inclusion of East German researchers,
- a comparison between levels of inclusion of East German researchers and of a control group.

Comparing east German researchers to a control group is necessary for two reasons. First, the measurement of changes in the inclusion of East German researchers before and after unification might be confounded by general trends in structures of scientific communities at that time. Secondly, the Web of Science database changes considerably over time, and the change is likely to affect some of the indicators.

Case selection

Researchers from the field of semiconductor physics were selected in order to limit variance of research and communication practices. Some internal variance is retained by including both experimental and theoretical research.
GDR semiconductor physicists working at universities and the Academy of Sciences of the GDR were identified among authors who published at least one article in the subject category "Physics, Condensed Matter" in SCI journals between 1980 and 1990. Among these, I selected researchers who had the opportunity to continue their career at a publicly funded research organization in East Germany (few of them did) for at least eight years in order to make sure that changes in inclusion patterns could materialize. The control group consists of semiconductor physicists from English-speaking countries with a similar research age (determined by the year in which they received their

PhD).[2] The sample consists of experimentalists as well as theoreticians and includes researchers from different universities in the US, UK and Ireland (table 2).

Table 2. *Overview over investigated cases*

Cases	Number of cases	Research age	Epistemic practice	Research organization in the 1980s
East German cases	10	PhD between 1968-1982	7 experimentalists, 3 theoreticians	5 at universities, 5 at the Academy of Sciences
Control group cases	10	PhD between 1968-1984	7 experimentalists, 3 theoreticians	from different universities in the US, UK and Ireland

Data

Meta data of researchers' publication oeuvre (including all publication categories such as journal articles, meeting abstracts, book chapters, etc.) between 1980 and 2000 were downloaded from the Web of Science. Metadata included titles, journal titles, authors, addresses, cited references and numbers of citations. Missing address information of a researcher's publications were manually added through searches of journal websites and full texts of articles. Data were further processed in Excel using VBA macros.

To avoid the shortcomings of the SCI Journal Impact Factor I calculated the visibility of those journals in which the 20 researchers published between 1980 and 2000 by normalizing the number of citations it received from other journal in the sample by the number of citable items. All journals that received more than 1000 citations in the first time period (1980-1990) respectively more than 2000 citations[3] in the second period (1991-2000) were categorized in tertiles of the normalized citations.

Results

Access to the community's published knowledge

In the GDR the supply with scientific literature from the West was limited (Gläser and Meske 1996: 330). The extent to which this restriction affected access to the community's knowledge might be reflected in reference lists (Table 3). I calculated the average length of reference lists and the proportion of references that were less than 3 years old for all articles by East German and control group researchers. In both periods, the East German physicists used more references than the control group. In the 1990s the number of references increases in both groups, which is likely caused

[2] For the control group, the database "ProQuest Dissertations & Theses Global" was searched for the keyword "semiconductor" in the thesis title and included the thesis supervisor as a member of the control group until 10 researchers were found. Including the authors of the theses themselves was impossible because most authors had left academia soon after receiving their PhD.

[3] The number of SCI publications in SCI journals has roughly doubled in the two time periods

by changing communication patterns worldwide.[4] Before German unification, East German physicists indeed cited fewer recent publications than physicists in the control group. This difference nearly vanished after German unification. The control group's former high share of references less than 3 years old (41,1%) requires further investigation.

Table 3. *Number and age of cited references in SCI journal articles*

Time span	*1980-1990*			*1991-2000*		
	References	*Mean number of references per article*	*References < 3 years*	*References*	*Mean number of references per article*	*References < 3 years*
East German group	3250	22,9	18.5%	9467	26,1	26,4%
Control group	7604	18,0	41.1%	11554	21,7	28,6%

To further compare the knowledge base of East German researchers to that of their Western colleagues, I identified the journals that were most frequently cited by the two groups (Table 4).

Table 4. *Dynamics of the most frequently cited journals in SCI publications by the two groups of physicists (East German journals in italics, shaded cells indicate journals present in both groups)*

Most frequently cited journals and number of occurrences in 1980-1990 by				*Most frequently cited journals and number of occurrences in 1991-2000 by*			
East German physicists		*Control group physicists*		*East German physicists*		*Control group physicists*	
Phys Rev B	450	Phys Rev B	923	Phys Rev B	2523	Phys Rev B	2156
Phys Status Solidi B	294	Appl Phys Lett	908	Phys Rev Lett	789	Phys Rev Lett	1643
Phys Status Solidi A	159	Phys Rev Lett	835	Surf Sci	329	Appl Phys Lett	1470
Phys Rev Lett	155	J Appl Phys	306	Appl Phys Lett	327	J Appl Phys	417
Phys Rev	153	Solid State Commun	290	J Appl Phys	311	Surf Sci	325
J Phys C Solid State	121	J Phys C Solid State	272	Solid State Commun	259	Opt Lett	317
J Chem Phys	118	Phys Rev	262	*Phys Status Solidi B*	256	J Vac Sci Technol B	241
J Cryst Growth	118	Surf Sci	261	Phys Rev	216	Phys Rev A	238
J Appl Phys	117	J Vac Sci Technol	251	Physica C	161	J Opt Soc Am B	232
Solid State Commun	94	Phys Rev A	169	J Cryst Growth	130	Semicond Sci Tech	160

[4] Differences in the number of references may also be caused by the distribution of articles across journals (some of which have restrictions concerning the length of articles or number of references).

The convergence of the two lists in the time after German unification is remarkable, with the first five journals being the same (marked in grey) and differences only occurring in the second half of the lists. These differences could be partly explained by researchers working in different research areas within semiconductor physics. An interesting difference is the high occurrence of references to articles in the East German journal "Physica Status Solidi". Although this journal was internationally recognized (Hoffmann 2013), it has been cited only 4 times in the 1980s by the control group physicists. In the 1990s, its importance for East German physicists decreased.

Offer of research contributions to the international community and their perception

For investigating East German researchers' inclusion in their communities through their scientific contributions, classical publication and citation indicators were used. Since the extent of inclusion varies considerably between researchers, I conducted this analysis at the individual level.[5] To compare different levels of publishing, I constructed tertiles of publication activity by calculating the mean number of publications and citations of all researchers (separately for the two time periods) and defining zero to two thirds of the mean as 'low', two thirds to four thirds of the mean as 'medium' and four thirds to twice the mean as 'high' publication activity. The same procedure was applied to citations the researchers received.

Among both East German and Western physicists, we find strong variation in publication and citation activities (Table 5). Some have been "silent", in Cole and Cole's (1967) categorisation, that is, they published little and were rarely cited. Among these, some remain silent in the second period (EG1-EG3, EG5, EG7, CG1). Three East German researchers (EG8-EG10) published regularly in the first period but were rarely cited. In contrast, we find more Western researchers who publish regularly and were frequently cited. Some of them (CG8 to CG10) could even be considered as "prolific" (Cole and Cole 1967) due to their many publications and citations. After unification, some East German researchers became more active in publishing (EG4-EG6), but the perception of their research remains low. Only one researcher (EG9), a theoretician, became a prolific physicist. The publication and citation dynamics also shows that the inclusion of researchers changes during their career and within research systems which have not undergone such radical changes such as East Germany.

If we consider citations as an indicator of visibility to the community (Gläser and Laudel 2001), it is important to explore to whom the East German researchers were visible. Who was and is citing these researchers? I coded the geographic location of citing publications according to the following categories: other East German physicists, colleagues from other countries of the Eastern Bloc, colleagues from West Germany, other Western colleagues, and Chinese colleagues (which are difficult to categorise as East or West). Multiple addresses from different communities were categorised as West Germany if at least one co-author belonged to this community, and as West if no West German but other Western addresses were present. A considerable number of addresses

[5] Applying bibliometric indicators at the individual level is notoriously problematic (Wouters et al. 2019). Individual-level indicators will not be interpreted as indicating performance and will be combined with other sources (archival records, interview data) in the study of inclusion wherever this is possible.

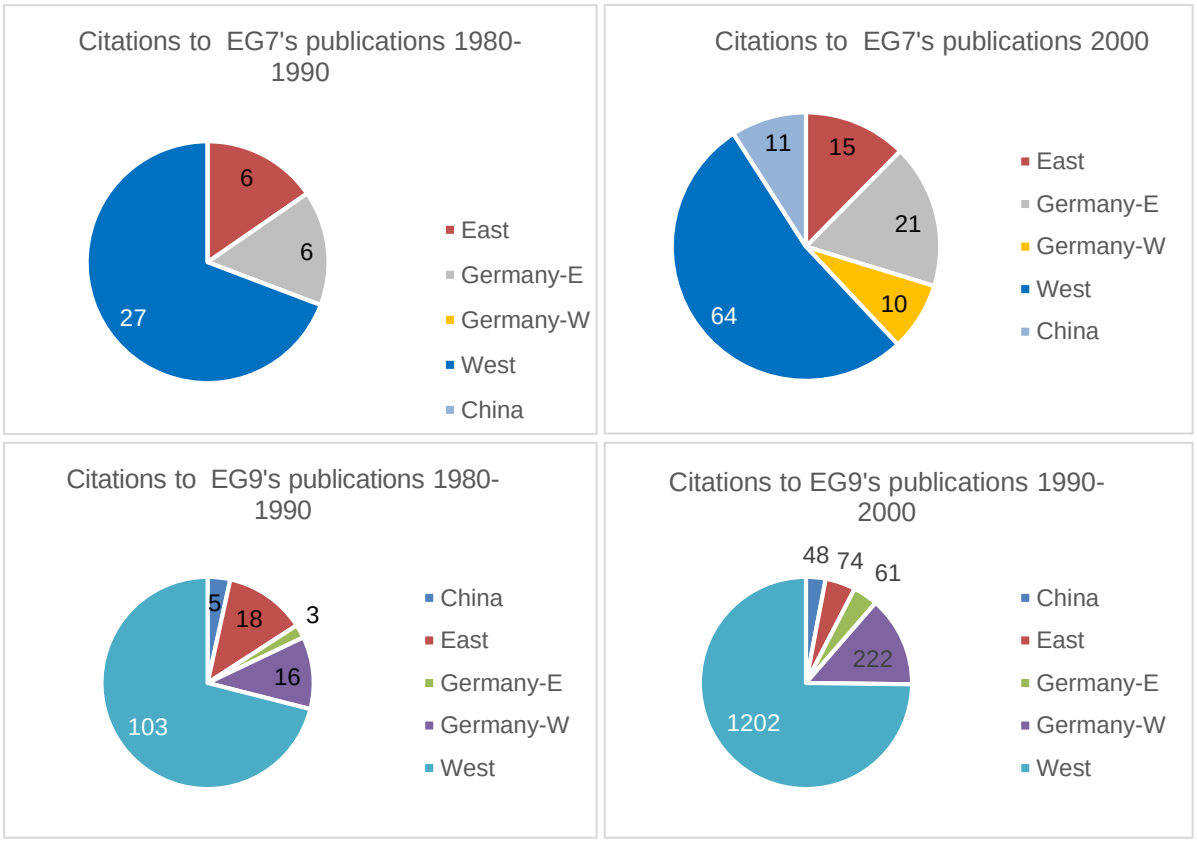
was missing (37% in the first period, 31% in the second period) which is particularly inconvenient in view of the low numbers of SCI publications of some East Germans. For this first test I assumed that the missing data would not change the distribution. Figure 1 shows the visibility of two East German researchers, one of which stayed "silent" (EG7) and one of which became "prolific"(EG9). Both researchers were predominantly cited by Western authors. This share increased after 1990 for the "prolific" researcher. His research contributions were also cited by West German researchers since the 1980s on.

Table 5. *Publication and citation dynamics of East German researchers (EG) and of the Western control group (CG)*

Case	*Publications 1980-90*		*Citations of Pub 1980-90 (until 94)*		*Publications 1991-2000*		*Citations of Pub 1991-2000 (until 04)*	
EG1	3	low	17	low	13	low	44	low
EG2	5	low	22	low	31	low	154	low
EG3	16	low	17	low	9	low	11	low
EG4	12	low	79	low	41	medium	121	low
EG5	14	low	70	low	15	low	70	low
EG6	21	low	66	low	56	medium	410	medium
EG7	23	low	39	low	29	low	121	low
EG8	36	medium	100	low	40	medium	127	low
EG9	37	medium	145	low	189	high	1607	high
EG10	38	medium	38	low	30	low	130	low
CG1	11	low	57	low	18	low	106	low
CG2	15	low	186	medium	80	high	1998	high
CG3	18	low	102	low	18	low	481	medium
CG4	26	medium	146	low	72	medium	527	medium
CG5	27	medium	179	medium	26	low	135	low
CG6	32	medium	280	medium	57	medium	208	low
CG7	34	medium	295	medium	2	low	10	low
CG8	51	high	414	high	40	medium	457	medium
CG9	110	high	828	high	259	high	1578	high
CG10	177	high	1789	high	119	high	1281	high

The analysis of the visibility of East German researchers in different regions of the world should be complemented by an analysis of their visibility to different strata of their community. For this, I analysed the visibility of researchers citing East Germans. Indeed, both researchers have been cited by authors whose publication received more than 100 citations.

Figure 1. *Two East German physicists' citing communities before and after 1990*



The chances of being visible do not only depend on the content of a researcher's contribution but also - and increasingly so - on the place where it is published. To explore the chances to be read by other researchers, I placed the journals (indicated in yellow) in which East German researchers published on two journal visibility maps (figures 2 and 3).[6] The inner circle contains the journals with the highest degree of visibility, the outer circle the journals with the lowest degree of visibility. This way, individual "journal footprints" of researchers can be constructed. If we look at two researchers' journal footprints, EG7 and EG4, we see that the journals in which they published before and after unification indeed became more visible. EG4 even dominantly published in the most visible journals.

[6] Proceedings papers and meeting abstracts were excluded because I was interested in a researcher's own publishing decisions and not those made by conference organizers.

Figure 2. *Visibility of journals in which East German physicist EG7 published before and after German unification (most frequently used journals circled)*

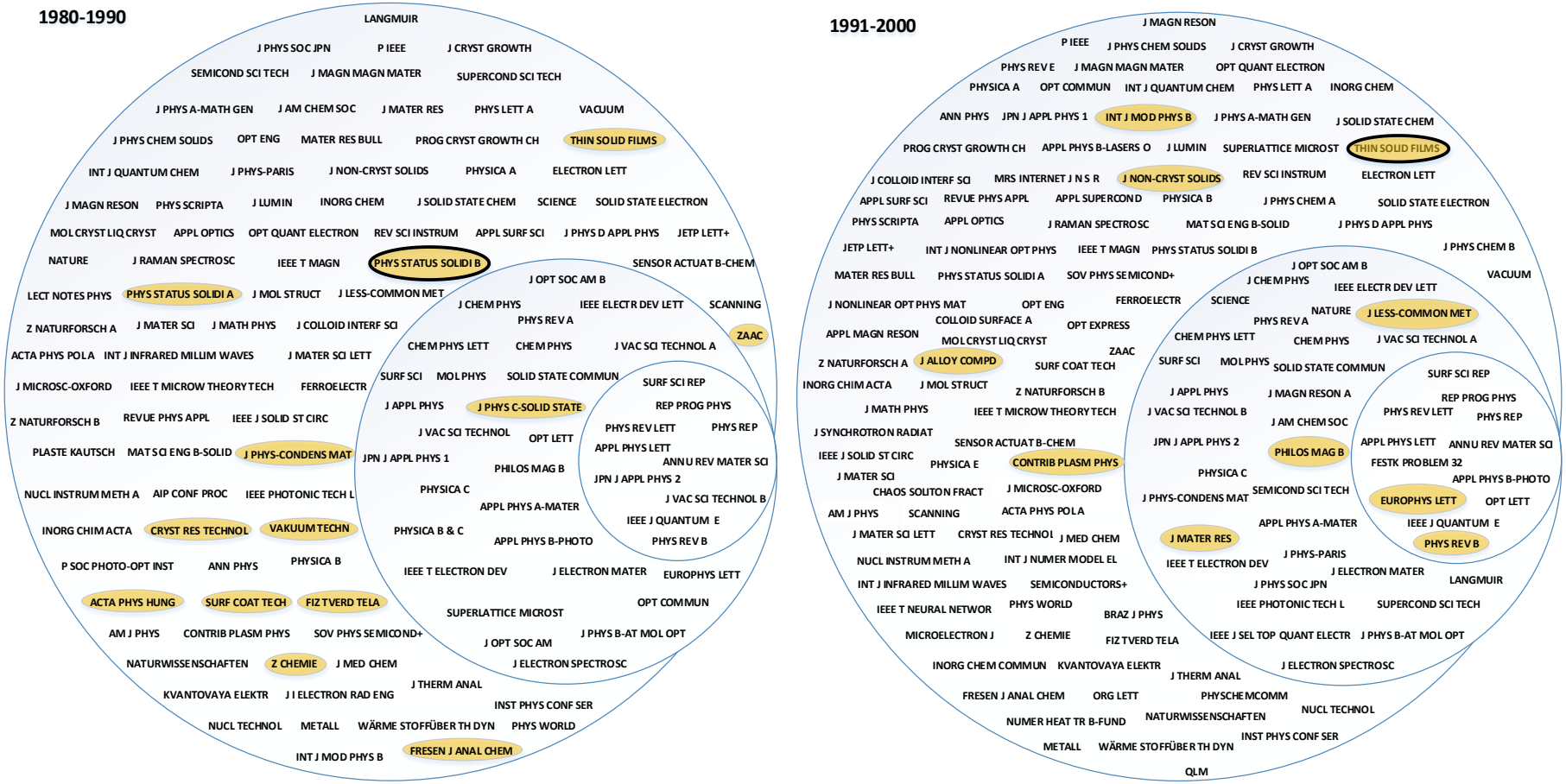
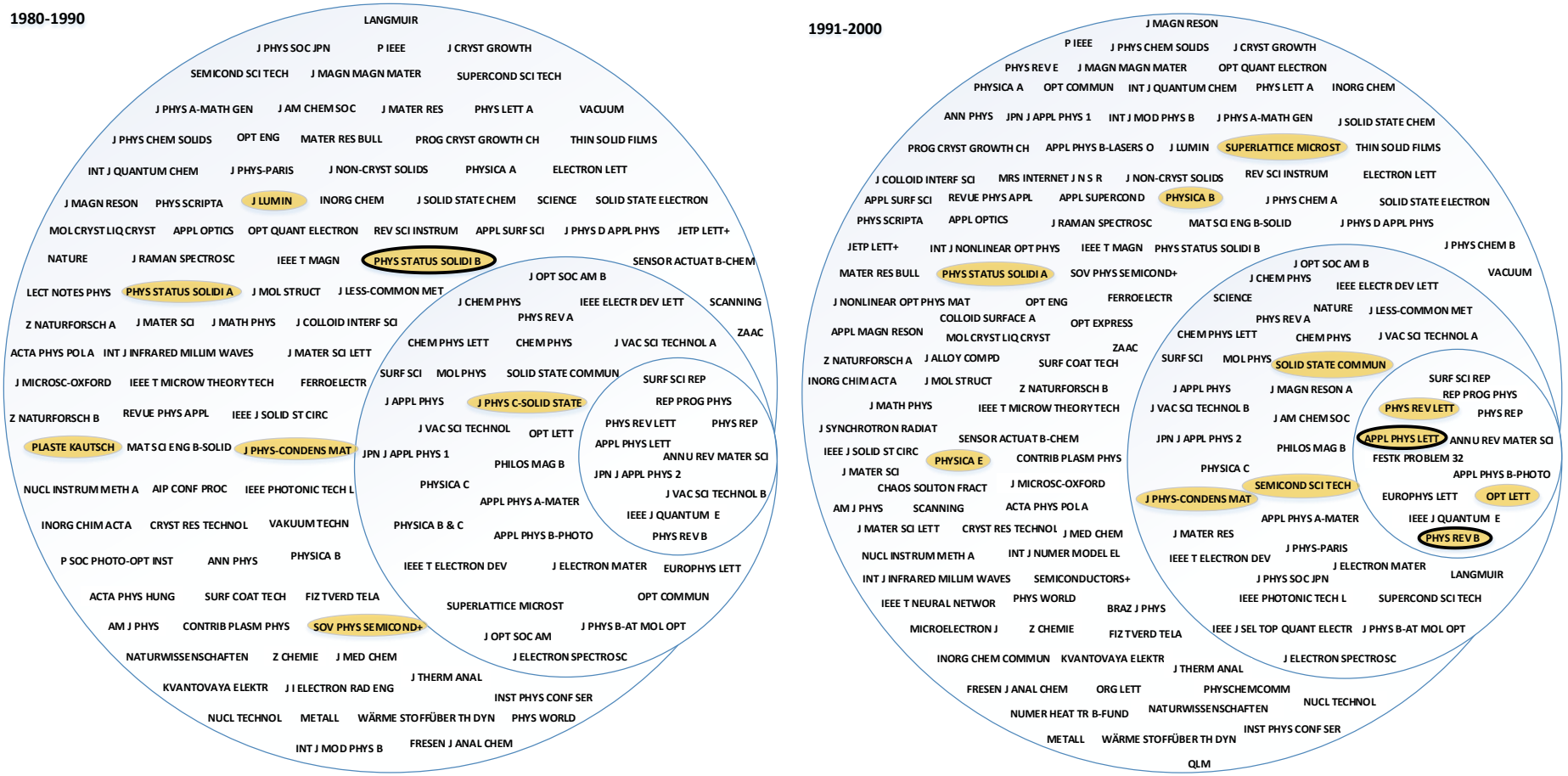


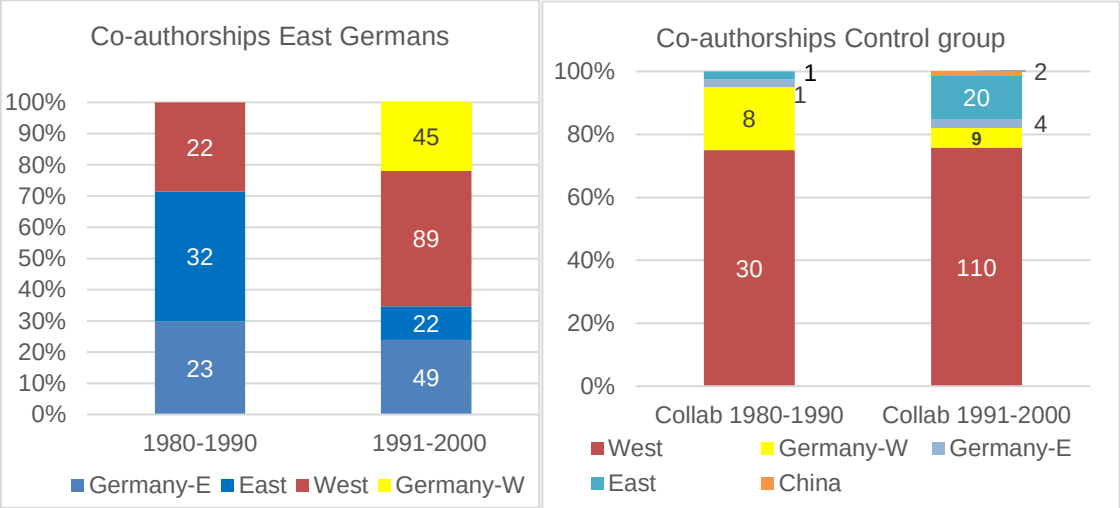
Figure 3. *Visibility of journals in which East German physicist EG4 published before and after German (most frequently used journals circled)*



Inclusion in collaborative research processes

The co-authorships of the 20 researchers were analyzed by categorizing the location of co-authors as described above. Figure 3 shows the collaboration patterns for the two groups. Before unification, East German physicists predominantly collaborated with researchers from the East. Half of them had collaborations with Western colleagues but nobody collaborated with colleagues from West Germany. The latter is not surprising because contacts to West German colleagues were particularly discouraged by the GDR government. Collaborations changed after German unification when collaborations with West German colleagues were no longer restricted. The control group researchers had hardly any collaborations with researchers from the East, a situation which changed after 1990.

Figure 3. *Development of co-authorships before and after 1990*



Conclusions

This test of bibliometric indicators demonstrates that it is possible to construct bibliometric inclusion profiles of researchers from meta-data of their publications, and that it is possible to observe differences in inclusion. The observed differences between individual inclusion profiles of East Germans strongly suggest that a centre-periphery perspective that puts whole national scientific communities either at the centre or in the periphery of scientific communities is insufficient for understanding inclusion. This is consistent with previous findings on research in the Global South (Davis and Eisemon 1989, Beigel 2014). A more differentiated analysis is not only required for theoretical reasons (inclusion is a characteristic of individuals and not of whole research systems) but it also makes it possible to use such profiles for further studies of causes and effects of inclusion.

Before sound conclusions about the suitability of bibliometric inclusion profiles can be drawn, the indicators must be validated with interview data, and indicators that best describe

inclusion and its change over time must be selected. Such a fine-grained approach makes possible the search for patterns of inclusion, whose identification in turn enables conclusions about inclusion patterns of national communities. For example, the small-scale analysis presented in this paper suggests that the existence of completely separate Eastern and Western sub-communities of the international semiconductor community in the 1970s and 1980s is unlikely. Inclusion appears to have always existed in some dimensions (at least for some researchers in some scientific communities such as semiconductor physics).

While bibliometric methods provide a unique approach to the measurement of inclusion, particularly from a historical perspective, they have two important limitations. First, the application of bibliometric indicators that utilise the WoS database must take into account changes in this database, with the number of journals, articles in journals, references in articles and citations continuously increasing due to changing communication practices and the changing coverage of the literature by the Web of Science. One way of doing this is the construction of control groups. The control group used here was still rather small and the influence of individual particularities and extreme values may have had distorting effects. A second problem is the limited coverage of WoS databases. Especially for the study of inclusion, all channels of formal communication must be investigated, including research publications not indexed in the SCI. Further analysis will take the research content and its dynamics into account, for example changes of topics as well as the content of collaborations. A third problem is the internal variance within the research field of semiconductor physics. These validity problems can be partly addressed by increasing the number of cases and by taking the internal cognitive structure of a community into consideration.

Acknowledgement

I am grateful to Ismael Rafols and Theresa Velden for their helpful comments on an earlier version of this paper. This work was supported by the German Ministry of Education and Research (Grant 01UJ1806CY).

References

Esquivel, A. V. and M. Rosvall (2011). "Compression of flow can reveal overlapping-module organization in networks." Physical Review X 1(2): 021025.

Arunachalam, S., and K. Manorama, 1988. How do Journals on the Periphery Compare with Mainstream Scientific Journals. Scientometrics 14: 83-95.

Beigel, Fernanda, 2014. Publishing from the periphery: Structural heterogeneity and segmented circuits. The evaluation of scientific publications for tenure in Argentina's CONICET. Current Sociology 62: 743-765.

Chavarro, Diego, Puay Tang, and Ismael Ràfols, 2017. Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling. Research Policy 46: 1666-1680.

Cole, Stephen, and Jonathan R. Cole, 1967. Scientific Output and Recognition, a Study in the Operation of the Reward System in Science. American Sociological Review 32: 377-390.

Davis, Ch. H.; Eisemon, Th. O., 1989. Mainstream and Non Mainstream Scientific Literature in Four Peripheral Asian Scientific communities. Scientometrics 15: 215-239.

Gläser, Jochen, and Grit Laudel, 2001. Integrating scientometric indicators into sociological studies: methodical and methodological problems. Scientometrics 52: 411-434.

Gläser, Jochen, and Werner Meske, 1996. Anwendungsorientierung von Grundlagenforschung? Erfahrungen der Akademie der Wissenschaften der DDR. Frankfurt a.M.: Campus.

Hoffmann, Dieter, 2013. Fifty years of physica status solidi in historical perspective. Physica Status Solidi B 250: 871-887.

Kuhn, Thomas, 1970. The Structure of Scientific Revolutions. Chicago: The University of Chicago Press.

Laudel, Grit, 2002. What do we measure by co-authorships? Research Evaluation 11: 3-15.

Merton, Robert K., 1973 [1938]. Science and the Social Order. Robert K. Merton (ed.), The Sociology of Science. Chicago: The University of Chicago Press, 254-266.

Polanyi, Michael, 1962. The Republic of Science. Minerva 1: 54-73.

Tijssen, Robert J. W., 2007. Africa's contribution to the worldwide research literature: New analytical perspectives, trends, and performance indicators. Scientometrics 71: 303-327.

Wouters, P., W. Glänzel, J. Gläser and I. Rafols (2013). "The dilemmas of performance indicators of individual researchers – An urgent debate in bibliometrics." ISSI Newsletter 9(3): 48 - 53.

Topic Reconstruction from Networks of Papers may not be possible if only one Algorithm is applied to only one Data Model

Matthias Held[1], Grit Laudel[2], Jochen Gläser[3]

[1] *matthias.held@tu-berlin.de, TU Berlin, HBS 7, Hardenbergstr. 16-18, 10623 Berlin (Germany)*
[2] *grit.laudel@tu-berlin.de, TU Berlin, FH 9-1, Fraunhoferstr. 33-36, 10587 Berlin (Germany)*
[3] *jochen.glaeser@tu-berlin.de, TU Berlin, HBS 7, Hardenbergstr. 16-18, 10623 Berlin (Germany)*

Introduction

The reconstruction of research topics from networks of papers is considered a major challenge that keeps attracting attention and for which new solutions are suggested (Šubelj et al. 2016, Gläser et al. 2017, Klavans and Boyack 2017, Held and Velden 2019). An interesting commonality of all approaches to topic reconstruction is that they are based on one data model, to which one algorithm is applied whose parameters are changed until it produces a 'satisfactory' solution. The same applies to experiments that systematically compare data models or algorithms. Whatever the data model or algorithm: The outcome of a topic reconstruction exercise is always a clustering solution that is produced by applying one algorithm with one particular setting of parameters to one data model.

The implicit assumption underlying this uniform strategy – that we just have to find the 'right' combination of data model and algorithm to obtain an accurate reconstruction of topics – remained unchallenged so far because testing it would require a 'ground truth' against which the outcomes of topic reconstruction exercises could be tested. If we consider the topics that orient researchers' work and which they shape with their publications as ground truths, then we have to acknowledge that there are multiple ground truths that are extremely difficult to access because they are conventions emerging from the use of publications in research rather than simply being knowledge structures documented in publications. Bibliometrics usually circumvents the problem by using expert validation (for a critique, see Gläser 2020), or by utilising bibliometric surrogates of researchers' views (Klavans and Boyack 2017).

In this paper, we report experiments that utilised an allocation of publications to topics through qualitative research as ground truths. We use these ground truths to test two data models (direct citation and bibliographic coupling) with two algorithms (the Leiden algorithm and Infomap). It turns out that although researchers' topics are reconstructed by these algorithms, it is impossible to predict at which parameter settings of the algorithms this happens. Furthermore, all four combinations of algorithms and data models have difficulties to reconstruct all topics with one specified set of parameters. Our preliminary conclusion is that a) multiple data models, algorithms or parameter settings for algorithms are necessary

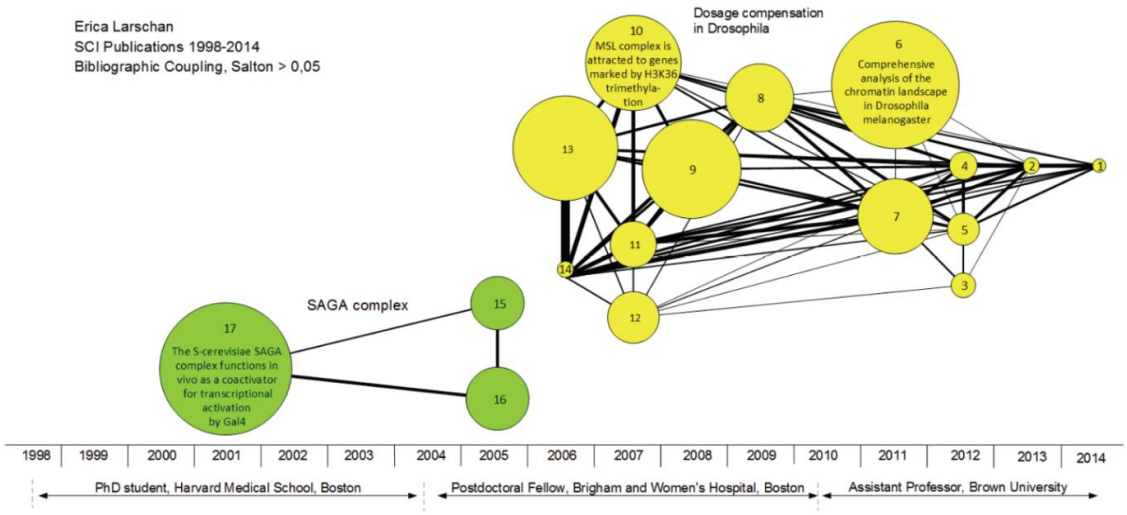
to reconstruct all or most topics in a set of papers, and b) that we are currently unable to predict which combination of data model, algorithm and parameter setting will adequately reconstruct which topics.

Data and Methods

Ground truths

The ground truths used in this project are topics of twelve physicists working in atomic and molecular optics who switched to experimental Bose-Einstein condensation (BEC) at different points in their careers. These topics were reconstructed in an internationally comparative project on conditions for scientific innovations, which included the emergence of experimental BEC in the 1990ies and 2000s in the field of Atomic and Molecular Optics (AMO) physics as one case (Laudel et al. 2014). The reconstruction of the BEC case was based on face-to-face semi-structured interviews whose main focus was the evolution of interviewees' research topics beginning with their PhD, with an emphasis on reasons for thematic change. To obtain the context necessary for understanding thematic changes, developments in the interviewee's national and international communities were discussed. To facilitate this part of the interviews, graphical representations of the researchers' research trails were constructed by downloading their publications from the Web of Science, constructing bibliographic coupling networks (using Salton's cosine for bibliographic coupling strength) and choosing a threshold for the strength of bibliographic coupling at which the network disaggregates into components (Gläser and Laudel 2015). While this 'manual' approach also produces several unassigned publications, it is preferable to algorithmic clustering because the research trails serve as means of 'graphic solicitation' in interviews, for which instant visual recognition of different topics is essential. The components represent topics a researcher has worked on over time (see Figure 1 for an example). The visualisations of individual research trails were used to prompt narratives about the content of the research at the beginning of the interview (for an extended description of the approach see Gläser and Laudel, 2015). During these narratives, researchers confirmed and sometimes corrected the picture by combining or separating clusters because they perceived research topics as belonging together or being separated. The interviews lasted on average 90 minutes and were fully transcribed. Transcripts were analysed by qualitative content analysis, i.e. we extracted relevant information from the transcripts by assigning it to categories that were derived from our conceptual framework (Gläser and Laudel 2013, Gläser and Laudel 2019).

Figure 1: *Picture of a researcher's cognitive career. It appears to consist of two research trails.*



Macro-level clustering

Data

To construct the macro-level AMO dataset, we first selected from the Web of Science all publications from journals in the subject category 'Physics, Atomic, Molecular & Chemical' published 1990-2005, excluding physical chemistry journals by excluding those carrying "chemi*" in the title. We then expanded this initial dataset by (1) including publications from all other physics subject categories (in the same time frame) that cited at least two publications from the journals belonging to the dataset in the time between 1975 and 2005; and (2) by including publications from all other physics subject categories which have been co-cited with at least two papers from our initial dataset. The direct citation network of this extended dataset has a giant component with 366,480 publications, which included all relevant publications of the research trails.

We obtained our macro-level AMO data set by applying the Leiden algorithm (see below, clustering) to this giant component and extracting the largest cluster, which contained 96,137 publications. This data set includes 80% of our twelve physicists' publications and 90% of these publications were assigned to topics. We used it to create two data models. The direct citation network was constructed by using only citations between publications in the dataset. Weights were attached to the links according to the formula given by Waltman and Eck (2012). The bibliographic coupling network was constructed using references that are source items in the Web of Science. Weights were attached to the links by calculating Salton's cosine.

Clustering

In order to detect communities in both networks, we selected two algorithms popular in the scientometrics community, namely the Leiden algorithm (an improved version of the Louvain algorithm), and the Infomap algorithm (Rosvall and Bergstrom 2008, Traag et al. 2018). Both algorithms include parameters that determine the resolution of the algorithm, i.e. the number

and sizes of clusters. The Leiden algorithm requires the specification of a *resolution parameter*. This parameter is included in its quality function CPM, which will be optimized for the chosen resolution value. The seed parameter of the Leiden algorithm was set to zero for all runs. Varying the resolution parameter leads to partitions with different numbers of clusters.
Infomap finds the minimum description length of a random walker in a given network by creating modules. The parameter *markov random time* has the standard value of 1. Changing this parameter means changing the number of steps of the random walker which are encoded (Kheirkhahzadeh et al. 2016). This will result in a more or less fine-grained solution. The intervals in which we varied the resolution parameter of the Leiden algorithm and markov random times of the Infomap algorithm are shown in Table 1.

Evaluation of Topic Reconstruction

In order to evaluate whether an individual researcher's research topic has been reconstructed by the macro-level clustering, we took the publications that were assigned to a topic by the researchers and assigned them to clusters in all cluster solutions. The assignment had to fulfil two criteria for qualifying as a successful topic reconstruction. (1) All publications of a topic belong to the same cluster, and (2) no publications from the researcher's other topics belong to this cluster. For the Infomap solutions, each of which has a hierarchical structure, we assigned publications to the lowest level of the hierarchy.

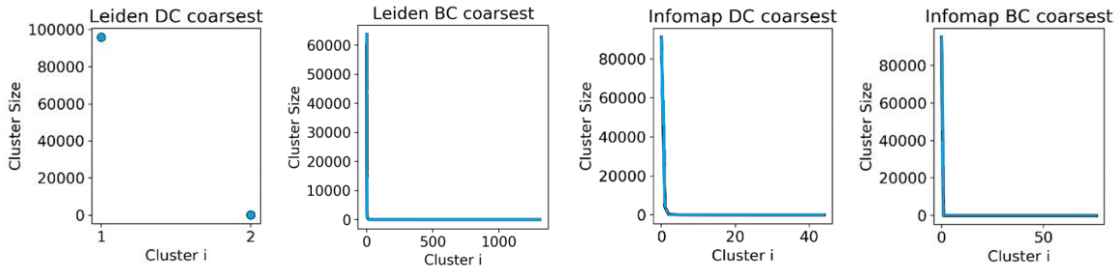
Results

Figure 3 displays the results for three of the twelve researchers and their topics. Each image shows at which level of granularity the topics were reconstructed (= 0, y-axis), were distributed over different clusters (= -1) or were combined with other topics in one cluster (= 1) by the specific combination of data model and algorithm. Each row corresponds to a topic and the four columns represent the results of the clustering solutions at different granularity levels. The x-axes correspond to the granularity levels also shown in Table 1 from coarsest to finest. Note that both algorithms – irrespective of the granularity level – create highly skewed cluster size distributions, i.e. few large clusters and very many small clusters. Figure 2 displays the distribution of cluster sizes at the coarsest levels.

***Table 1**. Numbers of clusters and sizes of smallest and largest cluster at smallest and largest granularity levels*

	Leiden DC		Leiden BC		Infomap DC		Infomap BC	
Resolution (Leiden) / mrt (Infomap)	min	max	min	max	min	max	min	max
	6e-07	5.44e-05	5e-05	1.36e-03	0.1	5.0	0.1	2.0
# Clusters	2	175	1,311	1,512	93,349	45	800	77
Cluster sizes (largest/smallest)	95,884 / 253	3,275 / 3	65,756 / 1	11,729 / 1	4 / 1	91,184 / 3	6,024 / 1	95,118 / 2

Figure 2: *Distribution of cluster sizes each of the lowest level of granularity*



The results for all three researchers have in common, that the assignments of topics to clusters are highly inconsistent (results for the other nine researchers are not different).

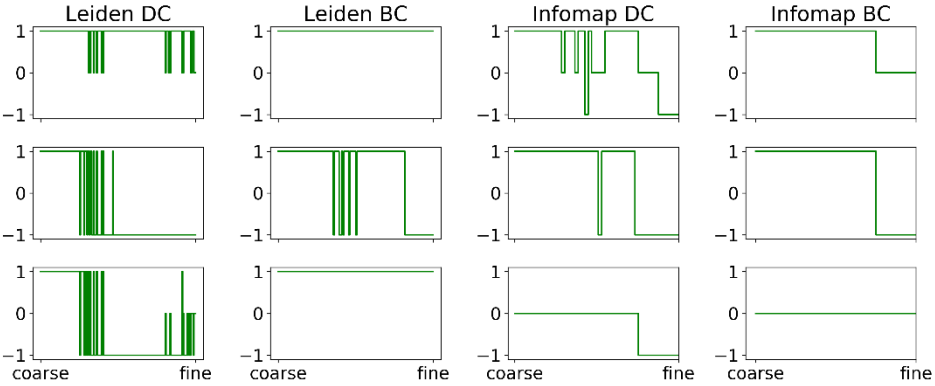
In general, the following patterns can be observed in Figure 3 when going from coarse to fine in a plot:

(i) We observe that topics are rarely reconstructed by a clustering solution. In most cases, topics are lumped together in clusters or spread across clusters.
(ii) In many cases, a topic is not reconstructed at any granularity level.
(iii) There are many fluctuations across granularity levels, i.e. publications belonging to one topic separate and combine again.
(iv) No clustering solution reconstructs all topics of a researcher.

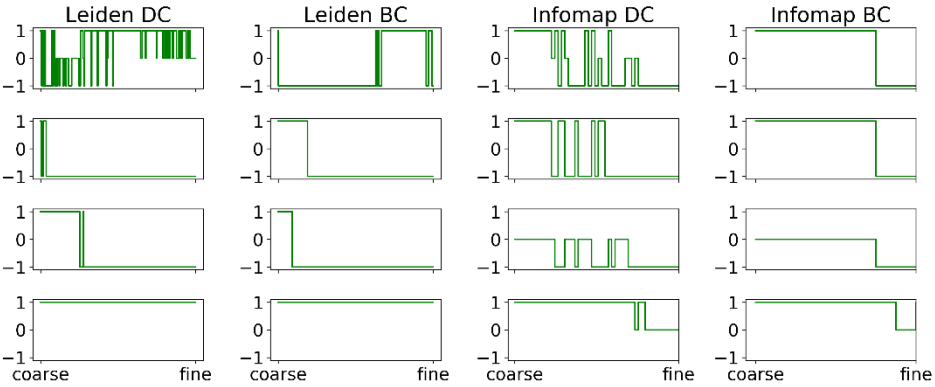
Thus, probably counterintuitively, there is no ’threshold’ granularity level above which the publications of a researcher’s topic are always split and assigned to different clusters. Instead, publications belonging to one topic are split at one granularity level and recombined again at a higher level or are lumped together with publications from one topic at a higher granularity level and then separated again. We refrain from a more detailed interpretation of this particular result because the two data models cannot be straightforwardly compared in terms of the fluctuations.

Figure 3: *Reconstruction of three researchers' individual topics by two algorithms applied to two data models (0: topic is reconstructed, -1: topic is were distributed over several clusters, 1: topic is combined with other topics in one cluster)*

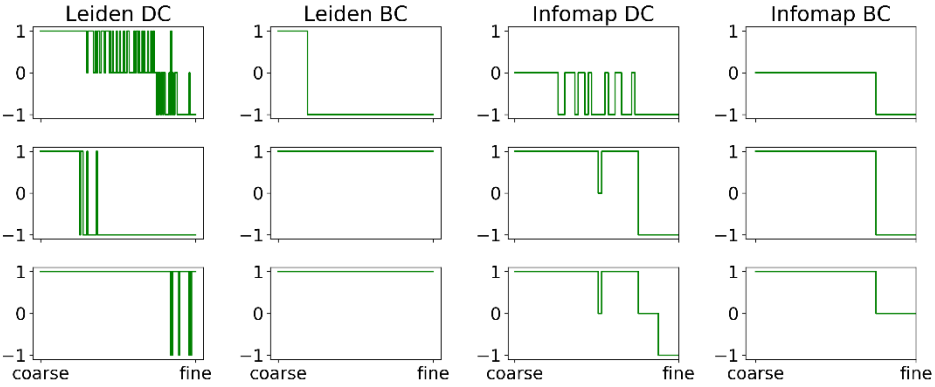
Researcher 1 (Three Topics)



Researcher 2 (Four Topics)



Researcher 3 (Three Topics)



Discussion

No combination of algorithm and data model could reconstruct all topics of one researcher at one level of granularity. The results indicate that each solution produced by the two algorithms at a particular granularity level either reconstructs a ground truth that remains unknown to us or produces *some* clusters representing *part* of our ground truth and others that must be considered artifacts. When compared with our researchers' topics, the results are highly inconsistent, i.e. topics are reconstructed unexpectedly, and we cannot know beforehand which topic is reconstructed by which combination of data model, algorithm and granularity level.

Why is this a problem? Differences between the perspectives of individual researchers on their topics and the macro-level cluster solutions, which in both data models include perspectives of thousands of researchers, were to be expected. However, what is unexpected is that we have the same inconsistent results for all 12 AMO researchers. If we define topics as shared perspectives on knowledge (Havemann et al. 2017: 1091), then the 12 researchers together not only represent a collective view on at least one shared topic (BEC), but also share their perspective on topics with other authors in our sample. These perspectives on the same topics might not be entirely congruent, but the inconsistency for all 12 perspectives appears to be significant. The topics of the twelve researchers represent one of the ground truths, and the reconstruction of even parts of that ground truths is unpredictable for all four data model/algorithm combinations across relevant resolution levels. Even after the exercise we cannot provide a single resolution level for any of the four combinations of data models and algorithms at which all topics are simultaneously reconstructed.

We conducted a preliminary assessment to find reasons on the micro level for the oscillation of publications of a researcher's topic between clusters at different levels of granularity. We found no correlation to (a) a weak bibliographic coupling between the publications, (b) an unusually high number of citations to publications, or (c) the content of oscillating topics (the topics being methods or theories).
Is this a problem of the algorithms? Both algorithms have in common, among other characteristics, that, first, they solely use the network topology, i.e. nodes (publications) and their edges (citation or coupling links) to detect community structures in the network. However, it does not yet seem entirely clear if topics are to be found in the topology (Gläser et al. 2017: 983). Secondly, both algorithms assign each node to exactly one community[7]. We know, however, that topics overlap in publications. So, when an algorithm performs a hard clustering on a publication network, the algorithms will be forced to assign publications to

[7] Note that there exists also a variant of Infomap that allows for overlapping communities, see Esquivel and Rosvall (2011). Specifically, this variant allows for boundary nodes - which have been determined before by Infomap to be assigned to more than one community.

clusters based on the best fit regardless of how good this fit is in absolute terms. This is why inconsistencies are to be expected.

Assuming that theoretically one should be able to detect topics in the topology of a direct citation or bibliographic coupling network, one can have a more specific look on the characteristics of the algorithms applied. Both algorithms, like all community detection algorithms, have built-in assumptions on what a community is and how it should be detected in a network. E.g., both optimize a function that is used for the entire network and assign every single node to a community. For the Leiden algorithm with the CPM quality function, a community is a set of nodes with a link density that is higher inside than between the other sets. For Infomap, a community is a set of nodes where each node can be reached from the others easily with only a few steps. Whether these assumptions are useful for the topic reconstruction endeavor we are undertaking is highly questionable in the light of our results. Furthermore, the mesostructures in networks which we assume to represent topics are difficult to find with an algorithm that is based on just one definition of community because we expect topics to have all kinds of structures and sizes (Havemann et al. 2017).

Conclusions

Our main conclusion is that since each topic reconstruction exercise simultaneously produces accurate and inaccurate representations of topics, achieving a valid reconstruction of topics through the combination of one data model and one algorithm with one specified setting of parameters is likely to be impossible. It might be necessary to combine several data models and algorithms to achieve a valid reconstruction. This depends on knowledge about the properties a topic must have in order to be reconstructed with a particular combination of data model and algorithm. This knowledge does not yet exist. Combining micro-level and content-based analyses with macro-level experiments of topic reconstruction might be a way forward towards identifying the variegated properties of topics and their connections to data models and algorithms.

Development and evaluation of topic reconstruction approaches need to proceed in parallel with enlarging the body of knowledge on topic properties in the various scientific disciplines. We need to answer questions like: How do topics form and develop in a particular field? Which bibliometric traces do these developments leave? Simply proceeding to develop and apply new or existing algorithms to reconstruct topics without any attempt at validation will not contribute to a cumulative knowledge creation in our field.

Acknowledgement

This research was supported by the German Federal Ministry of Education and Research (Grant 01PU17003).

References

Esquivel, A. V. and M. Rosvall (2011). "Compression of flow can reveal overlapping-module organization in networks." Physical Review X 1(2): 021025.

Gläser, J., W. Glänzel and A. Scharnhorst (2017). "Same data—different results? Towards a comparative approach to the identification of thematic structures in science." Scientometrics 111(2): 981-998.

Gläser, J. and G. Laudel (2013). "Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations [96 paragraphs]." Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research 14(2): Art. 5.

Gläser, J. and G. Laudel (2015). "A Bibliometric Reconstruction of Research Trails for Qualitative Investigations of Scientific Innovations." Historical Social Research / Historische Sozialforschung Vol. 40, No. 3 (2015): Special Issue: Methods of Innovation Research: Qualitative, Quantitative and Mixed Methods Approaches.

Gläser, J. and G. Laudel (2019). "The discovery of causal mechanisms: Extractive qualitative content analysis as a tool for process tracing [76 paragraphs]." Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research 20(3): Art. 29.

Havemann, F., J. Gläser and M. Heinz (2017). "Memetic search for overlapping topics based on a local evaluation of link communities." Scientometrics 111(2): 1089-1118.

Held, M. and T. Velden (2019). "How to interpret algorithmically constructed topical structures of research specialties? A case study comparing an internal and an external mapping of the topical structure of invasion biology." 10.

Kheirkhahzadeh, M., A. Lancichinetti and M. Rosvall (2016). "Efficient community detection of network flows for varying Markov times and bipartite networks." Physical Review E 93(3): 032309.

Klavans, R. and K. W. Boyack (2017). "Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge?" Journal of the Association for Information Science and Technology 68(4): 984-998.

Laudel, G., E. Lettkemann, R. Ramuz, L. Wedlin and R. Woolley (2014). Cold Atoms – Hot Research: High Risks, High Rewards in Five Different Authority Structures. Organizational Transformation And Scientific Change: The Impact Of Institutional Restructuring On Universities And Intellectual Innovation. R. Whitley and J. Gläser. Bingley, Emerald Group: 203-234.

Rosvall, M. and C. T. Bergstrom (2008). "Maps of random walks on complex networks reveal community structure." Proceedings of the National Academy of Sciences 105(4): 1118-1123.

Šubelj, L., N. J. van Eck and L. Waltman (2016). "Clustering Scientific Publications Based on Citation Relations: A Systematic Comparison of Different Methods." PLOS ONE 11(4): e0154404.

Traag, V., L. Waltman and N. J. van Eck (2018). "From Louvain to Leiden: guaranteeing well-connected communities." arXiv:1810.08473 [physics].

Waltman, L. and N. J. van Eck (2012). "A new methodology for constructing a publication-level classification system of science: A New Methodology for Constructing a Publication-Level Classification System of Science." Journal of the American Society for Information Science and Technology 63(12): 2378-2392.

Opening the Black Box of Expert Validation of Bibliometric Maps

Jochen Gläser

jochen.glaeser@tu-berlin.de, TU Berlin, HBS 7, Hardenbergstr.16-18, 10623 Berlin (Germany)

Introduction

One of the continuing worries of scholars engaged in bibliometric topic reconstruction is the question of these reconstructions' validity. Do we reconstruct the topics researchers work on? Shortly after the introduction of bibliometric mapping in the 1970s, these worries have initiated attempts to validate the maps, i.e. to identify a 'ground truth' (the topics researchers work on) and to assess how well this ground truth is reconstructed by mapping exercises (Gläser et al. 2017: 985-986).

More than three decades later, we are still worried, not least because we have not yet found a reliable method for validating bibliometric maps. Interest in validation has lapsed or been diverted to bibliometric surrogates of ground truths, with the concomitant move from validity to accuracy (Klavans and Boyack 2017).

The starting point of this paper is the observation that there has never been a proper methodological discussion of how to conduct a validation of maps. With few exceptions, validation exercises have been conducted ad hoc, with a marked imbalance between the sophistication of map-making and the sophistication of map-validation. As a consequence, the validation process has largely remained a black box.

In this paper, I open this black box by identifying methodological problems inherent to the validation of maps by experts and proposing a strategy that overcomes these problems. I begin by critically examining the state of the art concerning the validation of bibliometric maps (2). The problems discussed in the bibliometric literature point to some fundamental properties of map assessment as a process of social construction (3). The strategy of validating maps needs to take these properties into account (4).

State of the art

The current state of the art on the validation of bibliometric maps is by and large the one produced between the 1980s and 2000s, with the most thorough discussion provided by Tijssen (1993). Since then, interest in expert validation of maps appears to be on the decline.
Three approaches to expert validation can be distinguished. The *use of auto-expertise* refers to bibliometricians mapping their own field and thus being able to assess the clusters. This approach is quite common (Small 1981, Waltman et al. 2010, Havemann et al. 2012, Šubelj et al. 2016), mainly because its obvious advantage of technical and domain expertise being very well integrated. However, this approach does not avoid all problems of expert validation discussed below. Furthermore, it seems difficult to generalise findings on the accuracy of maps for one small field to all of science.

The validation of maps of fields that are different from that of the map maker sometimes turns to the *use of reflexive expertise of the domain*, i.e. of published reviews, textbooks, or histories of the field as a yardstick for the validation of maps (Nadel 1980). This approach must rely on the interpretation of accounts from domain experts by map makers.
The most common approach, which is the focus of this paper, is the *interactive validation of maps by domain experts*. Within this general approach, three strategies can be distinguished:

(1) Creating maps from expert knowledge and comparing them to bibliometric maps

McCain (1986) asked experts to sort cards with author names according to the similarity of these authors' research and used these similarities as input for multidimensional scaling. She compared the resulting map to her author-cocitation map. Peters and van Raan (1993) asked experts to supply words that characterised thematic developments and compared these words to their co-word maps. Tijssen (1993) let experts rate the strength of cognitive connections between topics represented by words, and subjected the resulting matrices to multidimensional scaling.

(2) Obtaining experts' opinions on the validity of bibliometric maps

Several researchers presented their maps to researchers from the mapped disciplines and asked them to assess the validity of the maps (Nadel 1980, Healey et al. 1986, Law et al. 1988, Tijssen 1993, Schwechheimer and Winterhager 2001). Although this practice was considered a validation 'method', authors provided very little information on the actual procedure of their validation exercises. Most researchers mention "workshops" or "interviews" without further specifying what happened. Only Peters and van Raan (1993) listed the questions they sent to experts.

(3) Obtaining experts' opinions on the usefulness of bibliometric maps

This approach was usually selected when bibliometric maps are constructed for policy experts, who were invited to assess the usefulness of the maps for the purpose for which they

were requested, namely to aide decisions on the distribution of funding between fields of science (Healey et al. 1986). The most sophisticated approach so far has been used by Klavans et al. (2012). The authors created cards on which they listed phrases, titles of publications, and names of authors for each cluster of their map, which they assumed to represent a "research problem". Scientists and scientists-turned-administrators were then asked to use these cards according to their familiarity with the topic. For the research problems they consider themselves experts in, they were then asked to conduct an additional sorting, e.g. identifying institutional strengths or identifying 'hot' research topics. In the overwhelming number of cases, experts could easily use the representations of topics, which Klavans et al. interpreted as a sign of the map's validity.

The findings on expert validations provided by these three strategies were similar. An assessment of (most of) the maps as valid dominated but there was always some disagreement between expert assessments and maps as well as among experts. The only exception appears to be the exercise by Klavans et al., which, however, could only draw an indirect conclusion from the successful use of the map for practical purposes to its validity. Since these assessments varied little across mapping approaches and approaches to expert assessment, they seem to be produced by general features of expert assessment rather than specific properties of maps or validation procedures. The discussion of features and problems of expert assessment can be summarised as follows:

- researchers are only competent for some areas of the map but not for others (Tijssen, Peters and van Raan, Klavans et al.),
- researchers are influenced by the maps they see, i.e. their assessment is not independent (McCain, Peters and van Raan),
- researchers may be guided by interest and relevance rather than validity (Tijssen), and
- different researchers have different criteria for evaluating the map (Tijssen).

Some of these problems were addressed by methodologies for expert evaluation. For example, McCain, Peters and van Raan, and Tijssen obtained input from experts without showing them the bibliometric maps. However, the overall agreement seems to be that the above-listed features of expert judgment, which are responsible for the at best partial validation of maps by experts, cannot be overcome. In particular, quantifying expert assessments and using them as input for creating map which are then compared to bibliometric maps seems problematic because it multiplies the problem. If there is disagreement between the two kinds of maps: which map-making procedure is to blame?

Sociological perspectives on the assessment of bibliometric maps by experts

The whole idea of validating bibliometric maps of scientific fields hinges on the definition of 'topic'. If a topic is defined as a knowledge structure represented by a cluster, the problem of validation does not exist because maps consist of topics by definition. However, this approach to defining topics either disassociates bibliometric mapping from all users of maps, which

have to deal with topics-for-researchers rather than topics-for-map makers, or just transforms the validation problem into the question of how topics-for-map makers correspond to topics-for-researchers. If bibliometric mapping wants to be useful to the sociology of science, it needs to adopt a sociological definition of 'topic', e.g. as "a focus on theoretical, methodological or empirical knowledge that is shared by a number of researchers and thereby provides these researchers with a joint frame of reference for the formulation of problems, the selection of methods or objects, the organisation of empirical data, or the interpretation of data" (Havemann et al. 2017: 1091).

With this sociological definition of a topic in mind, we can now move to a sociological understanding of the validation of maps. From a sociology of science perspective, the assessment of maps by experts from the mapped fields is a process of social construction. This premise has two implications. The first implication is that researchers develop their assessment of a map in the process of understanding it. A common misunderstanding of assessment procedures is that people carry around mental yardsticks, which they apply if an assessment task arises. This is not the case. Tijssen (1993) showed that if researchers carry mental maps of their field at all, they don't use them in the assessment of bibliometric maps. Ethnographic observations of schoolteachers' grading practices and of editorial decision making on journal articles are not scientific measurement processes. Instead, assessment criteria are developed and applied in interaction with that which is to be assessed (Kalthoff 2013, Hirschauer 2015). The same applies to the assessment of bibliometric maps of science by domain experts. This means that when confronted with a bibliometric map, researchers will attempt to make sense of the map, and will attempt to 'repair' inconsistencies through interpretation. Unless they fail completely, this process is likely to lead to the common "yes, but …" statement.

A second implication is that researchers interpret maps from their individual scientific perspectives. The cognitive resources researchers can bring to bear on the assessment task include their non-scientific and scientific frames and values, their scientific and non-scientific knowledge (including knowledge about norms of their scientific communities and the societies they live in), and their scientific and non-scientific interests. Since researchers' scientific knowledge is most extensive for the topics they are currently working on and decreases for topics more distant from their current concerns, researchers inevitably have very uneven knowledge about different regions of the map. They are also likely to apply different scientific perspectives to different regions of the map and may have specific interests concerning these regions. For example, a research may consider one area relevant due to a similarity of research objects to their own and others because of methodological advances they want to benefit from. Finally, researchers have a variety of scientific and non-scientific interests, all of which shape the interpretations of maps even if researchers want to be objective.

From these premises follows that the most likely outcome of any traditional validation exercise is a set of statements of the type "this is a good map, but …", with the "buts" varying according to the perspective of the assessor. Thus, each assessor is declaring a map to be partially valid and partially invalid, with the concept 'valid' taking on the meaning of 'corresponding to their current scientific perspective'. Since the aim of validation exercises usually is to establish the correspondence of the whole map to the topics shared by one or more scientific communities, the contribution of such individual statements to a validation is limited.

If the assessment of a map is a process of social construction that involves properties of assessors that are difficult to measure, and whose impact on the assessment is difficult to understand, then it seems logical to turn to qualitative methods of the sociology of science to improve the process. This I will attempt in the following section.

Understanding experts' understanding of maps

The basic idea of the proposed approach is to 'calibrate' researchers before they assess a map, and to let them assess maps by explicitly applying their scientific perspective. 'Calibration' means obtaining knowledge about the researcher's scientific perspective and interests for the dual purpose of acquiring information necessary for the later interpretation of their statements about the map and making this perspective an explicit point of reference for the remainder of the interview.

The approach suggested here is based on the sociological method of qualitative, partially standardised interviews (Gläser and Laudel 2010). Partially standardised means that some of the questions to be asked are decided upon prior to the interview and used in all interviews, while their phrasing, additional questions, and answers to the questions emerge unconstrained during the interview process. This method has been specified for interviews with researchers as 'scientifically informed interviewing' (Laudel and Gläser 2007) that is supported by graphic representations of 'research trails' (Gläser and Laudel 2015).

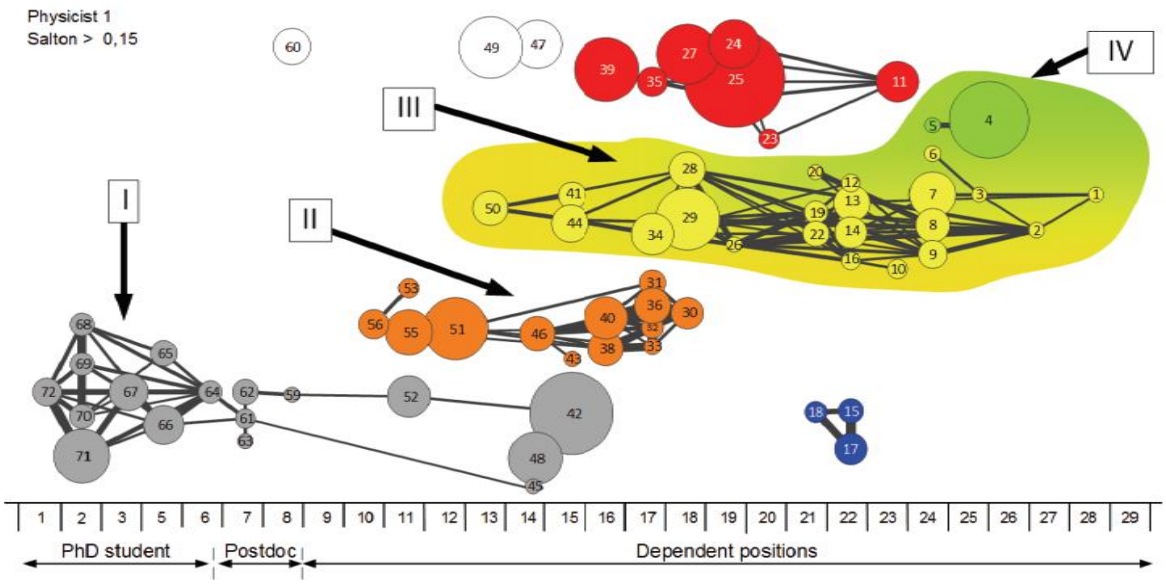
Reconstructing researchers' scientific perspectives

Interviews about maps should be conducted in two parts, with the first one devoted to eliciting a researcher's scientific perspective and the second part focusing on the application of that perspective in the researcher's assessment of the map. The scientific perspective of a researcher can be understood as an individual frame, i.e. as a taken-for granted cognitive scheme of an actor that provides knowledge about situations and proven solutions for typical problems (Schütz 1967, Schütz and Luckmann 1973, Goffman 1974). Frames guide actors' acquisition and evaluation of information without actors being fully aware of them. Frames can, however, be made explicit by discussing an actor's decisions and making them respond to challenges to their views. The particular frame we are interested in - a researcher's scientific

perspective – must be deduced from their prior and current research, the evolution of their interests and plans, and their reasoning about scientific decisions.

As an input for this first part of the interview, we construct the interviewee's research trail by performing a micro-level cluster analysis of the interviewee's oeuvre. This is done in a very simple way by constructing the bibliographic coupling network of the researcher's publications and raising the threshold for coupling strength until the network breaks down into components (Gläser and Laudel 2015, see Figure 1 for an example). Graphic representations of research trails are used in the interview to establish the sequence of topics a researcher has worked on, thematic changes and the reasons for them, and plans for further research. With this part of the interview, the regions of the map in which the researcher worked, further regions that are of actual or potential interest to the researcher, and scientific interests can be established.

Figure 1: *Graphical representation of a physicist's cognitive career, which includes five research trails of different size (source: Gläser and Laudel 2015: 315). Numbers in circles refer to a publication list. The width of lines indicates the strength of bibliographic coupling, the size of circles the numbers of citation received. Roman numerals were used to link sections of the map to the interview transcript.*



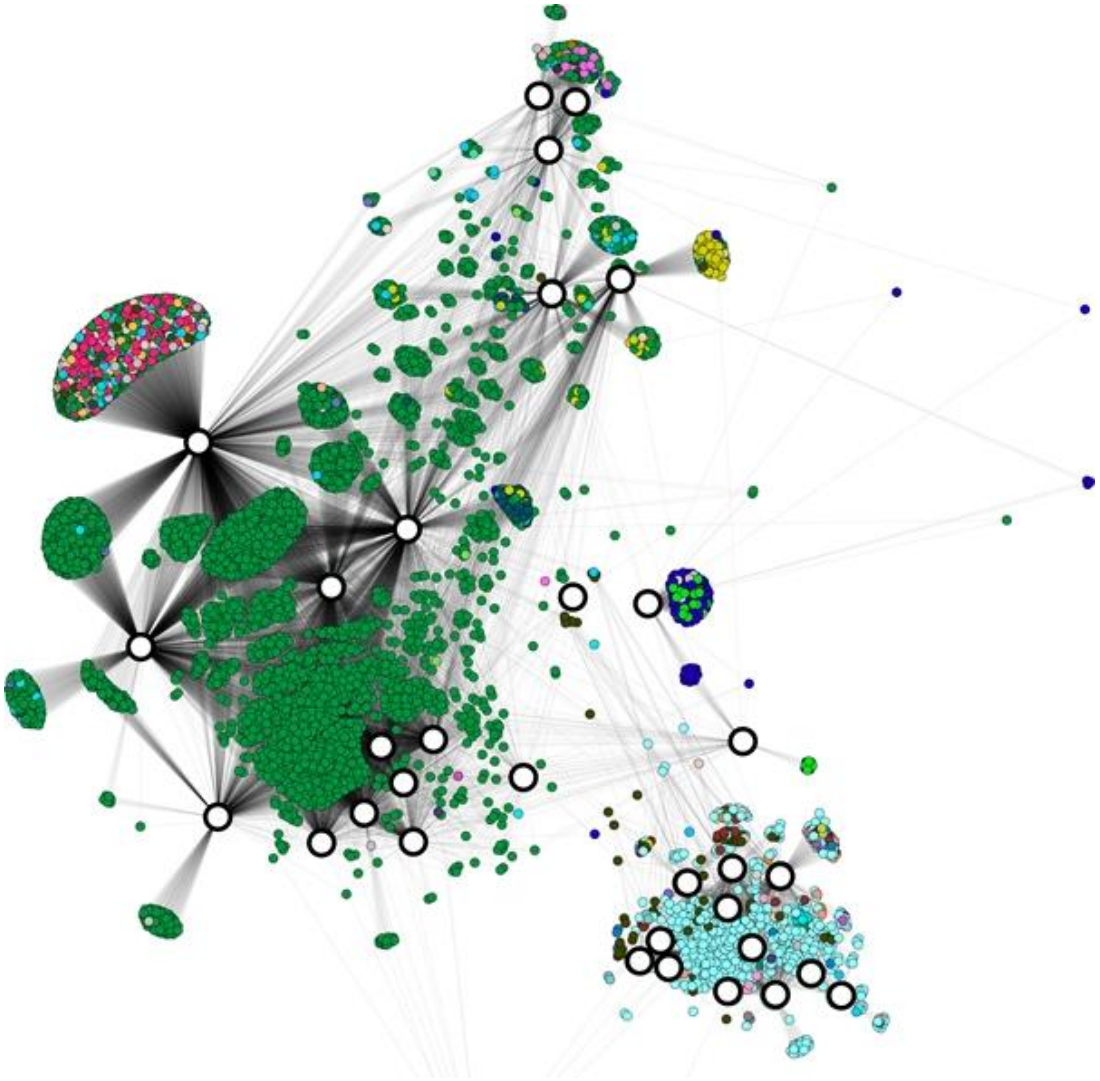
Discussing maps

On the basis of the preceding discussion of the interviewee's scientific perspective, the interviewee's assessment of the map can be obtained in the second part of the interview. This part begins by using the research trails to position the researcher on the map in order to provide an 'objective' starting point ('objective' in the sense that it is based on prior

publications rather than statements generated ad-hoc). Starting from this picture (see Figure 2 for a speculative example), all regions of the maps can be discussed with regard to

- the interviewee's knowledge about the research that is represented there,
- the scientific distance between the researcher's own work and the work allocated to this region,
- the interviewee's particular interest in the work represented by this region of the map, and
- colleagues who work on these topics and the evolution of their work.

Figure 2: *Example of a map in which the interviewee's publications (from Figure 1) are positioned (larger white circles with black borders).*



In this discussion, the parts of the map the interviewee can assess and the nature of the interviewee's knowledge about different regions of the map can be established. Experiences with discussions of research trails in interviews suggest that researchers will also readily point out regions of the map that don't make sense to them, should be split, or should be merged. Whenever the researcher indicates such disagreement between their perception of their field and the way in which it is represented by the map, the reasons for this disagreement can be explored. The focus of this exploration is be whether there could be a perspective from which the representation make sense, and how this perspective differs from that of the researcher.

Analysing interviews

Regardless of the specific method of data analysis employed, the main task of the analysis is the reconstruction of each interviewee's perspective that was used in discussing the map, the reconstruction of assessments of different regions of the map in the light of the scientific perspectives applied, and the aggregation of these assessment. In the analysis of each interview, the interviewee's relationship to each topic can be categorised. For topics the interviewee is engaged with, their position can be thematically central or marginal. For other topics they are outsiders with different degrees of ignorance and interest. For example, they can be beneficiaries, i.e. researchers using knowledge without contributing to the topic.

On this basis, statements about the topics can be interpreted and weighted when integrated. This should lead to an 'assessment map' that links information obtained form researchers and the trust that can be placed in this information to each topic, and thus to an overall assessment of the map.

Conclusions

With this paper I would like to suggest how the use of theoretical and methodological sociological knowledge can support a crucial task of the development of approaches to bibliometric topic reconstruction, namely the validation of maps (and through them, approaches) by experts from the mapped fields of research. The proposed approach is based on reconstructing experts' scientific perspectives and using these perspectives for the interpretation of their assessments of maps.

The proposed approach requires an extensive preparation of interviews, several one-on-one face-to-face interviews, the verbatim transcription of interviews and a careful in-depth analysis. If used, the validation of maps becomes a specialist task, and a collaboration between bibliometricians and sociologists is required. This is a substantial investment. Its advantage is that we can open the black box of expert validation and turn the validation exercise into a valid exercise.

Acknowledgement

I am indebted to Matthias Held for providing the map in Figure 2, and to Markus Hoffmann, Chris Grieser and Grit Laudel for helpful comments.

References

Gläser, J., W. Glänzel and A. Scharnhorst (2017). "Same data—different results? Towards a comparative approach to the identification of thematic structures in science." Scientometrics 111(2): 981-998.

Gläser, J. and G. Laudel (2010). Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen Wiesbaden, VS - Verlag für Sozialwissenschaften.

Gläser, J. and G. Laudel (2015). "A Bibliometric Reconstruction of Research Trails for Qualitative Investigations of Scientific Innovations." Historical Social Research / Historische Sozialforschung 40(3): 299-330.

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. Cambridge, Cambridge University Press.

Havemann, F., J. Gläser and M. Heinz (2017). "Memetic search for overlapping topics based on a local evaluation of link communities." Scientometrics 111(2): 1089-1118.

Havemann, F., J. Gläser, M. Heinz and A. Struck (2012). "Identifying Overlapping and Hierarchical Thematic Structures in Networks of Scholarly Papers: A Comparison of Three Approaches." *PLoS ONE* 7(3): e33255.

Healey, P., H. Rothman and P. K. Hoch (1986). "An experiment in science mapping for research planning." *Research Policy* 15(5): 233-251.

Hirschauer, S. (2015). "How Editors Decide. Oral Communication in Journal Peer Review." *Human Studies* 38(1): 37-55.

Kalthoff, H. (2013). "Practices of grading: an ethnographic study of educational assessment." *Ethnography and Education* 8(1): 89-104.

Klavans, R. and K. W. Boyack (2017). "Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge?" *Journal of the Association for Information Science and Technology* 68(4): 984-998.

Klavans, R., K. W. Boyack and H. Small (2012). *Indicators and precursors of "hot science."*. 17th International Conference on Science and Technology Indicators, Montreal, Canada.

Laudel, G. and J. Gläser (2007). "Interviewing Scientists." *Science, Technology & Innovation Studies* 3: 91-111.

Law, J., S. Bauin, J.-P. Courtial and J. Whittaker (1988). "Policy and the Mapping of Scientific Change: A Co-word Analysis of Research into Environmental Acidification." *Scientometrics* 14(3-4): 251-264.

McCain, K. W. (1986). "Cocited Author Mapping as a Valid Representation of Intellectual Structure." *Journal of the American Society for Information Science* 37(3): 111-122.

Nadel, E. (1980). "Multivariate Citation Analysis and the Changing Cognitive Organization in a Specialty of Physics." *Social Studies of Science* 10(4): 449-473.

Peters, B. (1993). *Die Integration moderner Gesellschaften*. Frankfurt a.M., Suhrkamp.

Peters, H. P. F. and A. F. J. van Raan (1993). "Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling." *Research Policy* 22(1): 23-45.

Schütz, A. (1967). *The Phenomenology of the Social World*. Evanston, Northwestern University Press.

Schütz, A. and T. Luckmann (1973). *The structures of the life-world*. Evanston, Northwestern University Press.

Schwechheimer, H. and M. Winterhager (2001). "Mapping interdisciplinary research fronts in neuroscience: A bibliometric view to retrograde amnesia." *Scientometrics* 51(1): 311-318.

Small, H. (1981). "The relationship of information science to the social sciences: A co-citation analysis." *Information Processing & Management* 17(1): 39-50.

Šubelj, L., N. J. van Eck and L. Waltman (2016). "Clustering Scientific Publications Based on Citation Relations: A Systematic Comparison of Different Methods." *PLoS ONE* 11(4): e0154404.

Tijssen, R. J. W. (1993). "A Scientometric Cognitive Study of Neural-Network Research: Expert Mental Maps Versus Bibliometric Maps." *Scientometrics* 28(1): 111-136.

Waltman, L., N. J. van Eck and E. C. M. Noyons (2010). "A unified approach to mapping and clustering of bibliometric networks." *Journal of Informetrics* 4(4): 629-635.

A Workflow for Creating Publication Databases from Scratch

Jenny Oltersdorf[1], Asja Mironenko[2] and Jochen Gläser[3]

[1] *jenny.oltersdorf@tu-berlin.de, TU Berlin, HBS 7, Hardenbergstr.16-18, 10623 Berlin (Germany*
[2] *a.mironenko@tu-berlin.de, TU Berlin, HBS 7, Hardenbergstr.16-18, 10623 Berlin (Germany)*
[3] *jochen.glaeser@tu-berlin.de, TU Berlin, HBS 7, Hardenbergstr.16-18, 10623 Berlin (Germany)*

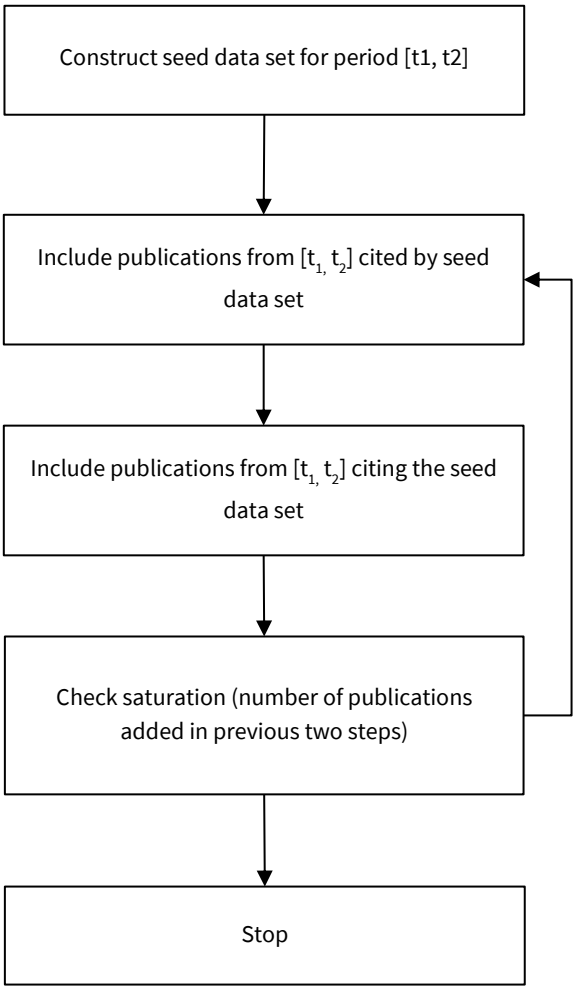
Introduction

In many cases, the utilization of bibliometric methods for solving research problems or practical problems depends on building a dedicated database because commercial databases such as the Web of Science (WoS) or Scopus do not sufficiently cover the literature of interest. In particular, researchers studying the social sciences and humanities often base their analyses on dedicated databases which they created manually (e.g. Ardanuy et al. 2009, Colavizza et al. 2018). In this extended poster abstract, we report part of our workflow for building such a database, which ideally should contain all German publications from a humanities field (Art History) and a social science field (International Relations) in a specific period. We will use the database to study the communication behaviour of the two fields through analyses of citation networks and citation contexts (Gläser and Oltersdorf 2019). Therefore, our dedicated database must include the publications' references and full texts (or at least the text surrounding the citations). The overall workflow for creating the database is based on the following algorithm (Figure 1):

1) Construction of a seed data set from publication lists of university academics.

2) Expansion of seed data by
   a. including publications from the target period that cite seed publications, and
   b. including publications from the target period that are cited by seed publications until saturation is reached.

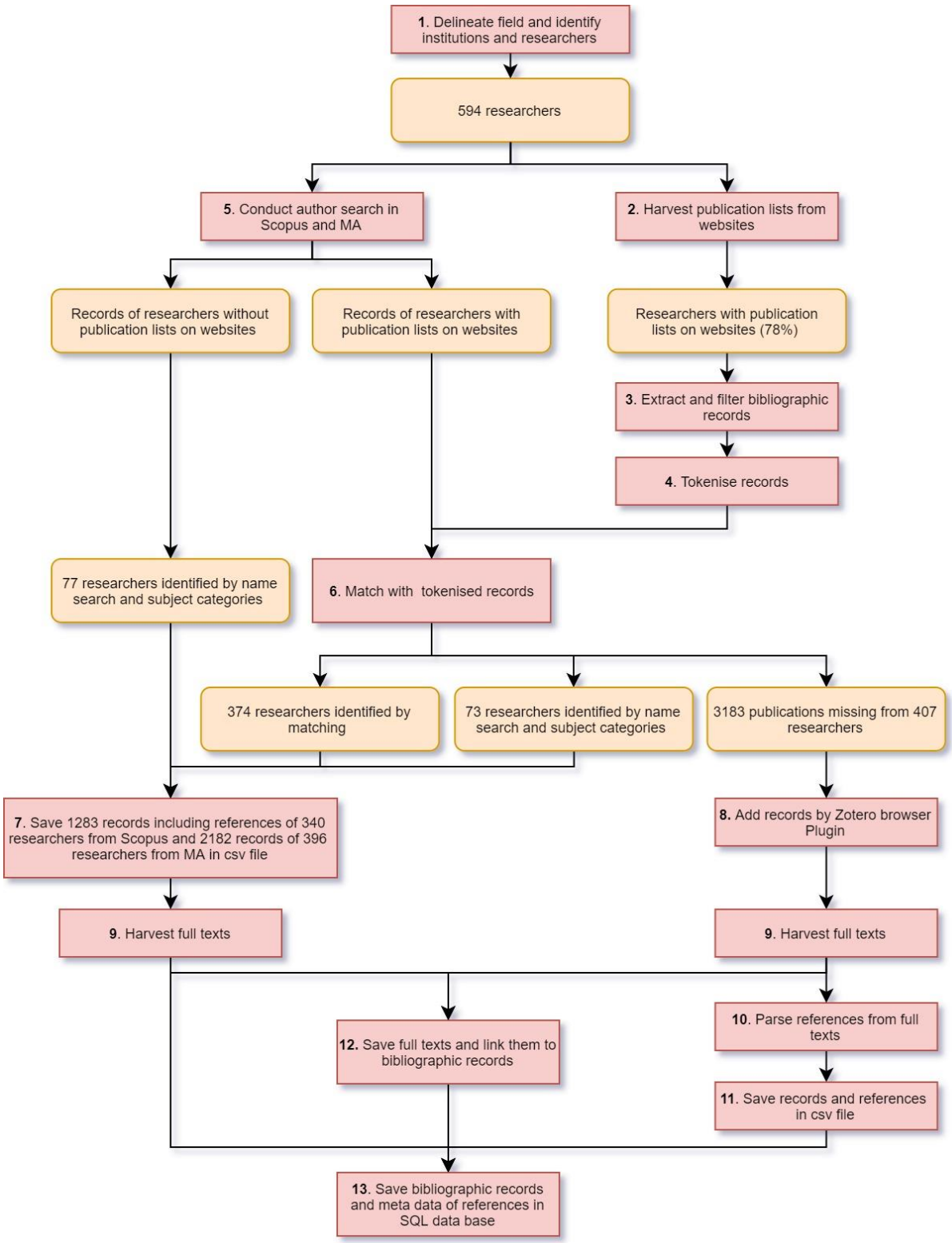
Figure 1: *Workflow for creating the database*



Workflow for creating the seed data set

In our extended poster abstract and poster we describe the first step, i.e. the construction of a seed data set, for the field of International Relations. The workflow includes the following steps (Figure 2):

Figure 2: *Detailed workflow for creating the seed data set*



1. Field delineation

Since the common approach to delineating a field by selecting publications does not work if no publication database with sufficient coverage is available, we first established criteria for the delineation of the two fields by consulting scholars in the field and library specialists, and by reading introductory texts. We decided to start from individuals and their affiliation to universities. For the field 'International Relations' (IR) we identified scholars who 1) have at least a master's degree and 2) are affiliated to a German university that 3) maintains a department with a research focus on IR. We disregarded nationality, country of graduation, and publication languages. This means that for creating the seed data set, we considered German IR to be represented by scholars who are currently located at German universities.
This decision could be challenged and merits further discussion. A closer look reveals that delineating a national sub-community is by no means trivial because nationality, publication language and geographic location do not coincide. In the case of Art History, the complexities are illustrated by the question whether German and non-German scholars working at the German Max-Planck-Institute 'Kunsthistorisches Institut in Florenz' in Florence or the German Max-Planck-Institute 'The Bibliotheca Hertziana' in Rome, who publish not only in German but also in other languages, should be considered as members of the German national Art History community.

2. Harvest publication lists from websites

We visited the university websites of all identified researchers in IR and used a Python script to convert publication lists on websites into PDF documents.[8] The automated conversion of publication lists was sometimes hampered by complicated website architectures (see Figure 3), access restrictions (see Figure 4), and other structural properties of websites. A total of 17% of the websites had to be converted manually. About 22% of researchers do not provide publication lists on their university websites or the publications are not in the period of interest. We conducted a search in Scopus and Microsoft Academic (MA) (see step 5) in order to include at least some of their publications. Further publications from these 22% of authors will be added with the expansion of the seed data set (see Figure 1).

[8] All scripts mentioned in this poster abstract were written by Asja Mironenko, and can be accessed on GitHub (https://github.com/mironenkoasja/bibliorecordsminer).

Figure 3: *Example of a website that defied automated conversion (problem: website architecture).*

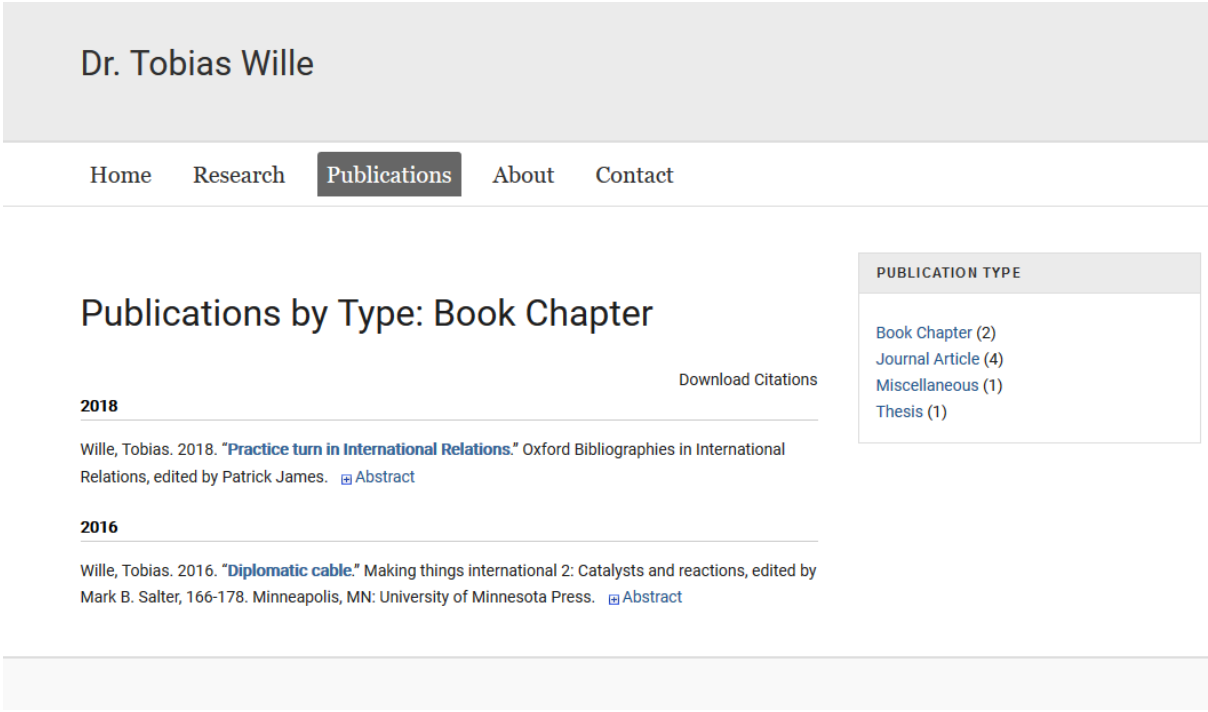
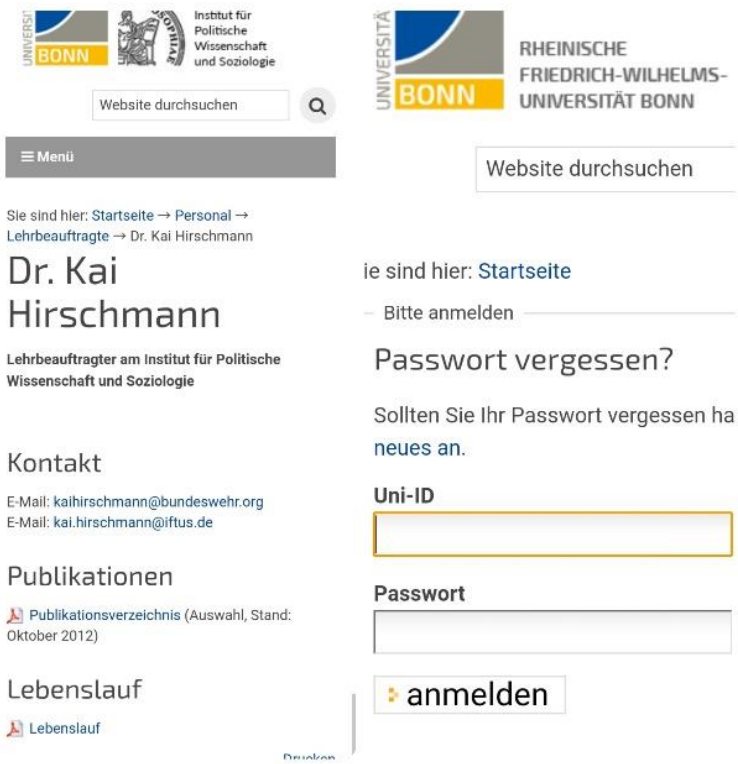


Figure 4: *Example of a website that defied automated conversion (problem: access only with password).*



3. Extract and filter bibliographic records

We extracted bibliographic records together with information about publication types from the PDF documents with a second Python script. The script was successful on 70% of the PDF documents. The extraction quality based on the Python script was determined by indicators of precision (0.996) and recall (0.957). The calculation was based on 403 records of 25 authors. Recall is the maximum percentage of records that can be identified. The precision rate is the percentage of the correctly extracted records in the period of interest that do not contain other text from the website such as biographical information or information on teaching experiences.
For the 30% of records that needed manual adjustment, the manual processing of a publication list took 6 minutes on average. Foreign-language titles increased the processing time significantly.

4. Tokenisation

The result of step 3 was a spreadsheet for each publication list that included bibliographic records as strings. We then tokenised the records, i.e. we split them into their bibliographic elements (author, title, year of publication, etc.). For this tokenisation we prepared the data by applying a third Python script, which removed irrelevant characters and built a consistent structure. The cleaned strings were processed using *AnyStyle*. *AnyStyle* is an open source parser for academic references which uses heuristics based on Conditional Random Fields for reference parsing.[9]
We assessed the performance of *AnyStyle* on the bibliographic elements author, title, and year by creating a scoring scheme. Each field was scored individually at 1 if the extraction was correct, at 0.5 if the element included additional characters, at 0.25 if characters were missing, and at 0 if the element could not be detected at all. The average quality for 100 randomly selected bibliographic records using AnyStyle 'as is' were 0.83 for author, 0.71 for title and 0.55 for year. After adjustment of AnyStyle's code the result for the year improved to 0.79. Even on the basis of partially incorrect tokenised records a matching algorithm could be applied (see step 6).

5. Conduct an author search in Scopus and Microsoft Academic

The two databases have the enormous advantage of providing tokenised bibliographic records, which in the case of Scopus include the references. This is why obtaining records from these databases for as many publications as possible is the most efficient approach. Retrieval of author names was conducted in the format 'lastname, initial of first name' in the data bases Scopus and Microsoft Academic (MA). We considered several name spelling alternatives if

[9] https://anystyle.io/

author names included umlauts or "*ß*". Umlauts were replaced by *ae and a*, *oe and o*, and *ue* and u, respectively because English-language journals and books use both versions of transliteration. The letter "*ß*" was replaced by "*s*" and "*ss*". Thus, the German Name Müßer would have to be searched for in the four variants Muser, Musser, Mueser, and Muesser.
A significant problem of the databases is homonyms. For authors who had publication lists on their websites, we eliminated homonyms by matching database entries with the tokenised records from harvested publication lists. If the results from the databases matched with a tokenised record, we considered the author as relevant and saved information on subject categories.
For the 22% of authors that had no publication list on their university websites, we only searched by author names and reduced the likelihood of homonyms by limiting results to the subject categories we derived from successfully matched records (see below, 6.).
Publications in MA are tagged "[…] with fields of study, or topics, using artificial intelligence and semantic understanding of content. Topics are organized in a non-mutually exclusive hierarchy with 19 top-level fields of study." (Microsoft 2020).[10] We searched for author names and filtered the results based on the most often used fields of study.

6. Match Scopus and MA records with tokenized records from publication lists

Matching was done by a fourth Python script, which utilises author names, publication titles, and publication years in an implementation of the Ratcliff/Obershelp pattern recognition algorithm (Black, 2014). The matching procedure identified 264 authors in Scopus. In addition, 47 authors with publications lists on their websites but no title match in Scopus and 29 authors without publication lists on their websites could be identified by an author-name search that was limited to relevant subject categories in order to exclude homonyms. Relevant subject categories were derived from the 264 successfully matched items and include categories like 'Political Science and International Relations', 'Social Sciences (all)' or 'Sociology and Political Science'. All in all, 340 or 57% of the 594 researchers are covered by Scopus (see Figure 6).
We used the same procedure in MA and identified 344 authors through matching, another 26 authors who provide a publication list and 48 authors who did not publish literature through a name search that was limited to relevant subject categories. In this process, 418 or 70% of researchers have been identified in MA (see Figure 7).
The overlap of authors indexed in Scopus and in MA can be found in Figure 8. Interestingly, the matching rate with MA varies considerably between publication types. We were able to match 70% of tokenised records for monographs, 58% of journal articles and 21% of working papers.

[10] Further information about the modeling process of fields of study can be found online https://www.microsoft.com/en-us/research/project/academic/articles/expanding-concept-understanding-in-microsoft-academic-graph/.

Figure 6: *Publication list - Scopus overlap*

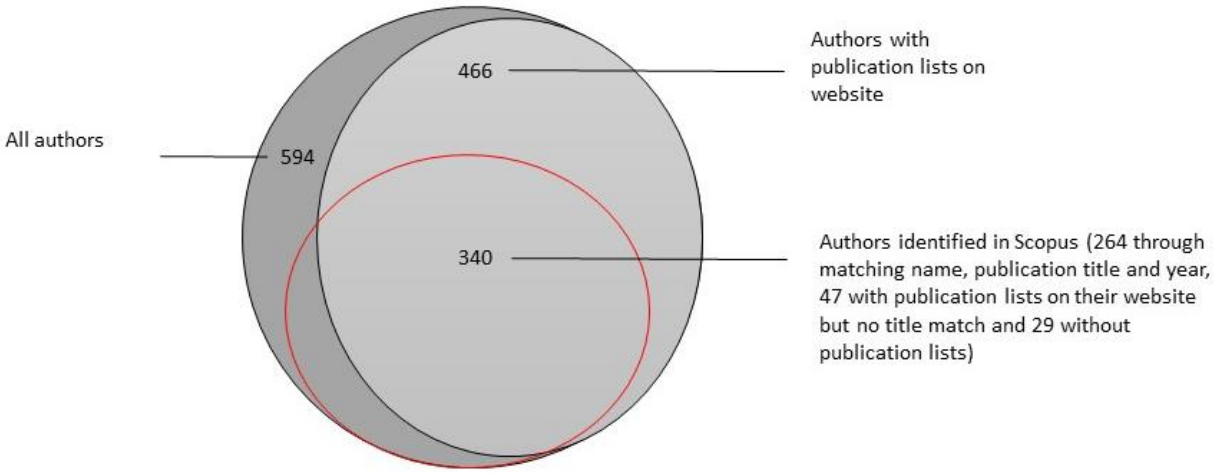


Figure 7: *Publication list - MA overlap*

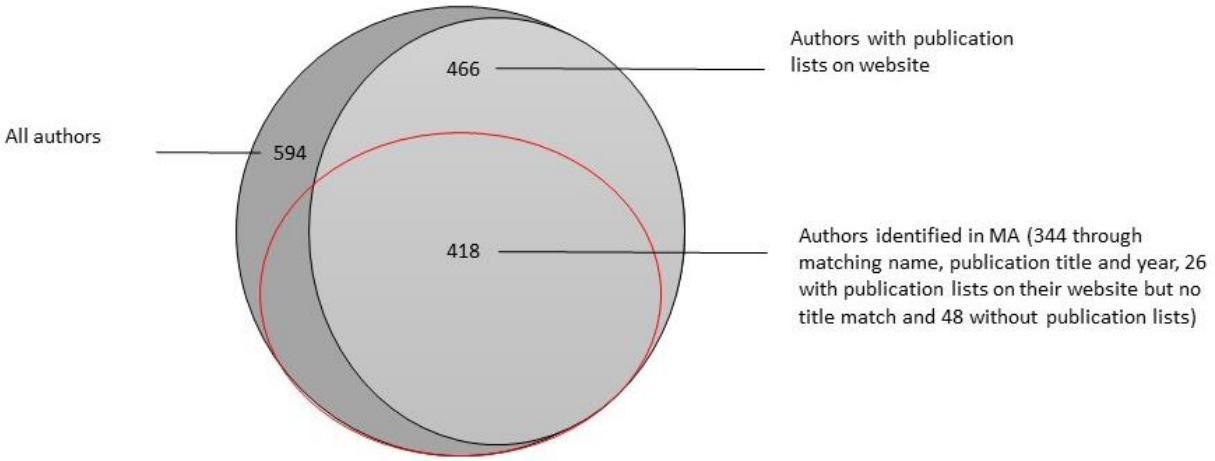
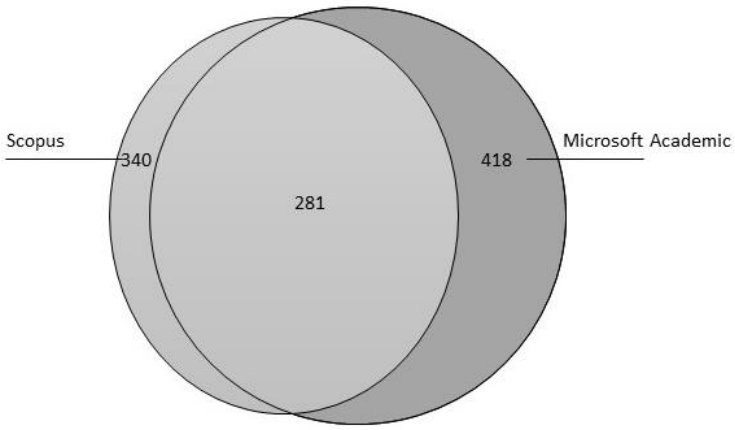


Figure 8: *Scopus - MA overlap*



7. Save records and references in csv format

From Scopus we saved 1293 records and references from 340 authors for further processing. In addition, we saved 2182 records from MA that are not in Scopus. In the case of MA, only records were saved as reference information is provided only in form of links to other items indexed in MA. We would have missed bibliographic information about MA's non-source items.

8. Add records by Zotero Browser Plugin

Those records that were only on the publication lists and could not be retrieved from the databases were stored in Zotero using the Zotero browser plugin *Connector*. Zotero is an open source reference management system[11] The Zotero Connector imports bibliographic meta data from, e.g., library catalogues or websites to the personal Zotero library. From there, data can be exported in several formats for further processing. The quality of imported data depends heavily on the used sources. Usually it varies among library catalogues, repositories, and websites. Therefore, all entries had to be checked manually and corrected if necessary.

9. Harvest full texts

Harvesting full texts was necessary for two reasons. First, many publications are not covered by Scopus, which also provides references. For these publications, references must be extracted from full texts. Secondly, full texts (or at least the citation contexts, which need to be extracted from full texts) are needed for the planned citation context analysis.
We combined several approaches to harvest full texts:

- We utilised information from publication lists on websites. Some researchers added links to full texts of at least some of their publications.
- We used the DOI, which is provided by MA and Scopus, to search for full texts.
- We placed national and international library loan requests to get hold of printed publications. We digitised the relevant parts (full texts of articles and chapters, references and citation contexts for books) for further processing.
- In a final step, we circulated letters to all researchers. We introduced the project and announced a follow up email. After two weeks, we sent them emails with the literature lists we created and asked if they would check for completeness, enhance the list if necessary, and provide us with the missing full texts that were labeled in the list. None of the scientists raised any objections to the procedure. So far, about 40% of them supported our project.

10. Parse references from full texts

Automatic reference detection in social sciences and humanities (SSH) publications still constitutes a major challenge. Most tools that extract references are based on referencing

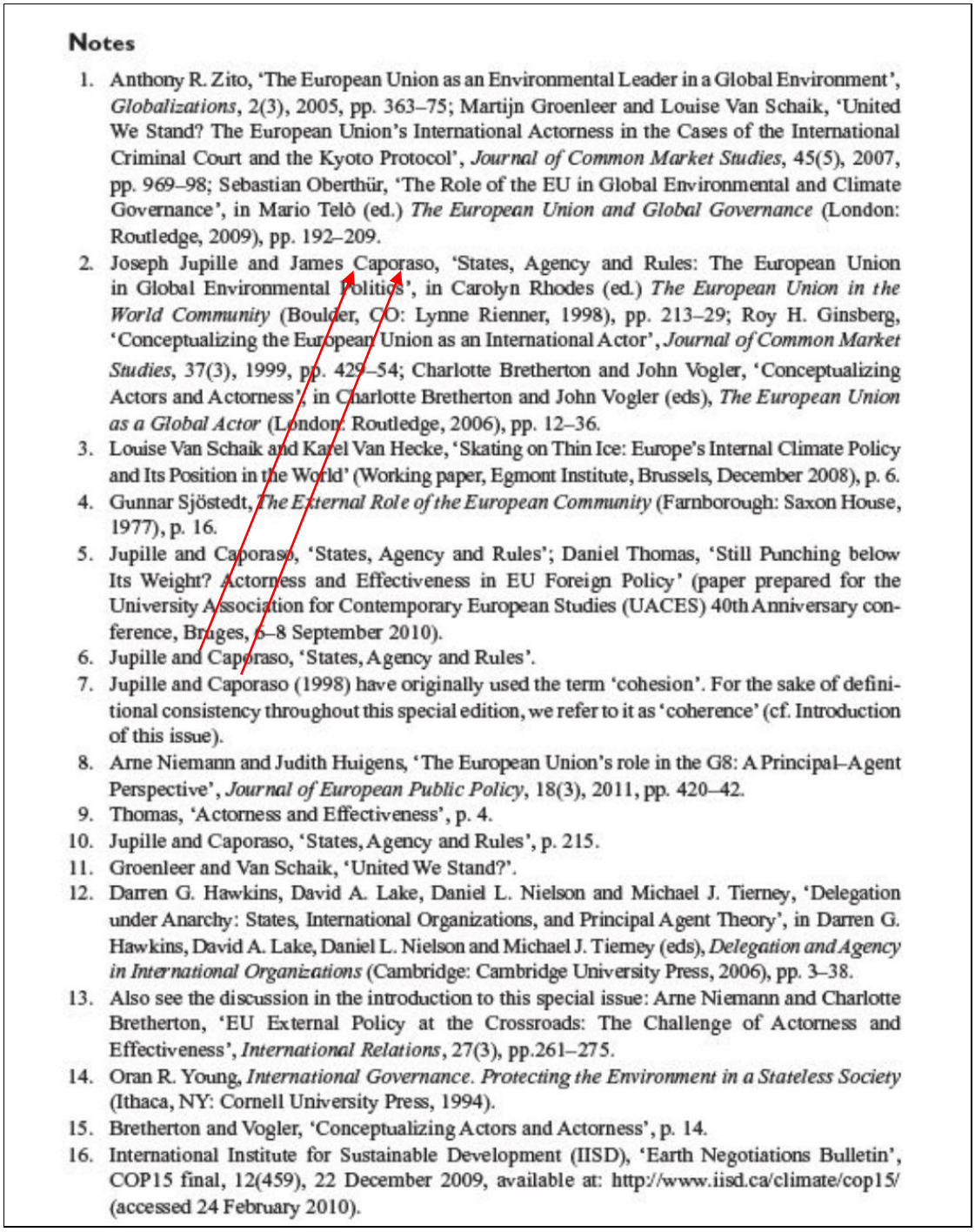
[11] https://www.zotero.org/

patterns of publications from the sciences and ignore the respective information for SSH publications. Furthermore, these tools cannot be applied directly to images, i.e. scanned documents. SSH publications that are not available in digital form must be scanned and the images must be processed with OCR software. This creates an additional source of errors which has a direct effect on the performance of the tools. To make things worse, the application of automated tools for reference extraction requires the existence of standardised reference sections at the end of a publication. In some SSH publications, however, full references are given in footnotes or side notes, and no reference section exists at the end of the publication.

For these reasons, none of the available tools met our requirements out of the box. A combination of tools with text-based and layout-based approaches in an iterative procedure that included manual adjustment turned out to be the best solution to extract references of publications in the field of International Relations.

First tests indicated that the problems multiply for the field of Art History. The publication behaviour of German Art History is still dominated by print publications. There is no consistent pattern of the positioning or phrasing of references across publications. For example, references may occur at the end of a publication in a separate reference list, in endnotes, footnotes, side notes, or insertions in the text. Bibliographic meta data in footnotes or endnotes are embedded into text (see Figure 9) or complex abbreviations are used to refer to other sources (see Figure 10). The current state of our investigations suggests that references cannot be extracted from publications in German Art History by any of the state-of-the-art tools such as CERMINE, GROBID, and ParsCit on its own. A combination of tools together and additional manual adjustments are needed. Another promising approach appears to be the layout-based approach to extract references from images that is based on complex neural networks (Rizvi et al., 2019).

Figure 9: *Reference section from journal 'International Relations' – A reference section where bibliographic information is first given as full bibliographic record and subsequently either as short titles or in the form "Author (year) without links to the full reference.*



Notes

1. Anthony R. Zito, 'The European Union as an Environmental Leader in a Global Environment', *Globalizations*, 2(3), 2005, pp. 363–75; Martijn Groenleer and Louise Van Schaik, 'United We Stand? The European Union's International Actorness in the Cases of the International Criminal Court and the Kyoto Protocol', *Journal of Common Market Studies*, 45(5), 2007, pp. 969–98; Sebastian Oberthür, 'The Role of the EU in Global Environmental and Climate Governance', in Mario Telò (ed.) *The European Union and Global Governance* (London: Routledge, 2009), pp. 192–209.
2. Joseph Jupille and James Caporaso, 'States, Agency and Rules: The European Union in Global Environmental Politics', in Carolyn Rhodes (ed.) *The European Union in the World Community* (Boulder, CO: Lynne Rienner, 1998), pp. 213–29; Roy H. Ginsberg, 'Conceptualizing the European Union as an International Actor', *Journal of Common Market Studies*, 37(3), 1999, pp. 429–54; Charlotte Bretherton and John Vogler, 'Conceptualizing Actors and Actorness', in Charlotte Bretherton and John Vogler (eds), *The European Union as a Global Actor* (London: Routledge, 2006), pp. 12–36.
3. Louise Van Schaik and Karel Van Hecke, 'Skating on Thin Ice: Europe's Internal Climate Policy and Its Position in the World' (Working paper, Egmont Institute, Brussels, December 2008), p. 6.
4. Gunnar Sjöstedt, *The External Role of the European Community* (Farnborough: Saxon House, 1977), p. 16.
5. Jupille and Caporaso, 'States, Agency and Rules'; Daniel Thomas, 'Still Punching below Its Weight? Actorness and Effectiveness in EU Foreign Policy' (paper prepared for the University Association for Contemporary European Studies (UACES) 40th Anniversary conference, Bruges, 6–8 September 2010).
6. Jupille and Caporaso, 'States, Agency and Rules'.
7. Jupille and Caporaso (1998) have originally used the term 'cohesion'. For the sake of definitional consistency throughout this special edition, we refer to it as 'coherence' (cf. Introduction of this issue).
8. Arne Niemann and Judith Huigens, 'The European Union's role in the G8: A Principal–Agent Perspective', *Journal of European Public Policy*, 18(3), 2011, pp. 420–42.
9. Thomas, 'Actorness and Effectiveness', p. 4.
10. Jupille and Caporaso, 'States, Agency and Rules', p. 215.
11. Groenleer and Van Schaik, 'United We Stand?'.
12. Darren G. Hawkins, David A. Lake, Daniel L. Nielson and Michael J. Tierney, 'Delegation under Anarchy: States, International Organizations, and Principal Agent Theory', in Darren G. Hawkins, David A. Lake, Daniel L. Nielson and Michael J. Tierney (eds), *Delegation and Agency in International Organizations* (Cambridge: Cambridge University Press, 2006), pp. 3–38.
13. Also see the discussion in the introduction to this special issue: Arne Niemann and Charlotte Bretherton, 'EU External Policy at the Crossroads: The Challenge of Actorness and Effectiveness', *International Relations*, 27(3), pp.261–275.
14. Oran R. Young, *International Governance. Protecting the Environment in a Stateless Society* (Ithaca, NY: Cornell University Press, 1994).
15. Bretherton and Vogler, 'Conceptualizing Actors and Actorness', p. 14.
16. International Institute for Sustainable Development (IISD), 'Earth Negotiations Bulletin', COP15 final, 12(459), 22 December 2009, available at: http://www.iisd.ca/climate/cop15/ (accessed 24 February 2010).

Figure 10: *Reference section from 'Allgemeines Künstlerlexikon' - Complex abbreviations that require external sources for decoding*

sonst (Entwurf und Ausführung des Kastells Montepoggiolo 1471) als Festungsbaumeister bezeugt. Die ihm früher zugeschriebenen Portale von S. Domenico (1481) und S. Agostino (1484) in Recanati werden jetzt wohl mit Recht als Arbeiten lombardischer Meister erachtet. Quellen: Vasari-Milanesi, Vite, II 467 ff; deutsche Ausg. Gottschewski u. Gronau, III 218 ff. - Gaye, Carteggio, I. C. v. FabriczyinJahrb.d.pr. Kunstsamml., XXIV (1903), Beih. p. 137 ff. (Regesten u. Urkunden.) - P. Gianuizzin Arch. stor. dell' arte, I (1888) 419 (Dom in Loreto, Urkunden). - C. v. Fabriczyin Repertoriumfür Kunstw. XX (1897) 87-96 (Tätigkeit in Neapel). - Serlio, Libri dell' architett., III. Bologna 1540 (Plan u. Beschreibg. von Poggioreale). - Colombo in Arch. stor. p. le prov. Napolet., X (1885) 186 ff:, 309 ff. (Poggioreale); IX (1884) 563ff. (Pal. d. Duchesca; Urkunden). - Ceci, eben da , XXIX (1904) 784ff. (Grabstätte). - G. Poggi in Riv. d'Arte, III (1905) (Neri di Bicci u. G., Urkunden). Allgemeines: Geymüller u.Stegmann, Architektur d. Renaiss. in Toskana, 1885-96, IV. - C. v. Fabriczy, Brunelleschi (Stuttgart 1895), passim. - Burekhardt, Cicerone, 11 120 f. - Patzak, Pal. u. Villa in Toskana, II (Leipzig 1913), passim; Villa Imperiale in Pesaro, 131 f. - Willich, Baukunst der Renaiss. in Ital. (Burger, Hdb. d. Kunstw.), p. 66 f. Einzelne Werke: G. Poggi, Duomo di Firenze, Ital. Forschungen, herausg. v. Kunsthist. Inst. in Florenz, II (1909) p. CXXII (Chorschranken des Doms).-Fri da Schott[üller, Amtl. Berichte aus d. Kgl. preuß. Kunstsancml., XXXIX (1917118) 79ff., (Intarsia und Schnitzarbeit; Türen). - C. v. Fahriczy, Jahrb. d. pr. Kunstsamml., XXIV (1903) 320 ff., (G. in Siena), ders., eben da , XXVI 40 ff. (G. in Macerata); ders. in Arch. stor. d. arte, III (1890) 441 ff. (Dom Faenza). - Supino, eben da , VI 154 f. (Dom in Pisa, Intarsien). - Papini, Cat. d. cose d'arte e d'antichità, Ser. I, Heft II, Pisa (Teil I, 1912). - Schmarsow, Melozzo da Forli (Stuttgart 1886) 122 (Loreto, Dom). - Bode, Denkmäler d. Renaiss.-Sculpt. Toskanas, 1892. - Rolfs, Franz Laurana, 1907 p. 238 ff. (Arbeiten in Neapel). - P. Frank I, Entwicklungsphasen d. neueren Baukst., Lpzg.1914 p. 39. - Folnesics, in Jahrb. d. Kunsthist. Inst. d. K. K. Zentralkom., VIII (1914) 87 (Portale in Recanati). - L. Venturi in L'Arte, XVII (1919) (Intarsien in Urbino). Max Semrau.

11. Save records and references in csv format

Records and references were saved in csv format for further processing.

12. Save full texts and link them to bibliographic records

To organise our data, we assigned identifiers to each author and each publication. Full texts were labelled with the publication ID and saved in separate folders for each author on an external hard disk.

13. Save bibliographic records and meta data of references in SQL data base

We set up an offline SQL database for further analysis in the project.

Concluding remarks

With this paper, we present for discussion a workflow for a type of project that has been occurring repeatedly in bibliometric research. Many theoretically interesting and politically important problems cannot be studied with commercial databases due to the latter's exclusion of publications from the Global South, from the Social Sciences and Humanities, and in languages other than English. So far, each scholar appears to have wrestled in isolation with the many practical problems involved in creating databases from scratch. We should discuss these practical problems in order to create a more efficient approach.

Acknowledgement

This research was supported by the German Federal Ministry of Education and Research (Grant 01PU17022). As our workflow demonstrates, building such a database as ours requires tireless and very precise manual labour, for which we thank our student assistants Elaheh Sadat Ahmadi, Liesa Houben, and Lisa Jura.

References

Ardanuy, J., C. Urbano und L. Quintana (2009). A citation analysis of Catalan literary studies (1974–2003): Towards a bibliometrics of humanities studies in minority languages. Scientometrics 81(2): 347.

Colavizza, G., M. Romanello and F. Kaplan (2018). "The references of references: a method to enrich humanities library catalogs with citation data." *International Journal on Digital Libraries* 19(2): 151-161.

Gläser, J. and J. Oltersdorf (2019). Persistent Problems for a Bibliometrics of Social Sciences and Humanities and How to Overcome Them. *Proceedings of the 17th INTERNATIONAL CONFERENCE ON SCIENTOMETRICS & INFORMETRICS*, Rome, September 2-5, 2019**:** 1056-1567.

Black, Paul E. "Ratcliff/Obershelp pattern recognition", in *Dictionary of Algorithms and Data Structures* [online], Paul E. Black, ed. 17 December 2004. (accessed TODAY) Available from: https://www.nist.gov/dads/HTML/ratcliffObershelp.html

Microsoft. 2020. „Microsoft Academic“. *Topics Analytics*. Abgerufen 25. August 2020 (https://academic.microsoft.com/topics).

Rizvi, S. T. R., Lucieri, A., Dengel, A., & Ahmed, S. (2019). Benchmarking Object Detection Networks for Image Based Reference Detection in Document Images. *2019 Digital Image Computing: Techniques and Applications (DICTA)*, 1–8. https://doi.org/10.1109/DICTA47822.2019.8945991