Homogeneous and heterogeneous scientific collaboration in Berlin metropolitan region

DEKiF case study 4

Aliakbar Akbaritabar (Ali)

Akbaritabar@DZHW.eu

21 January, 2020

DZHW
German Centre for Higher Education Research and Science Studies

Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

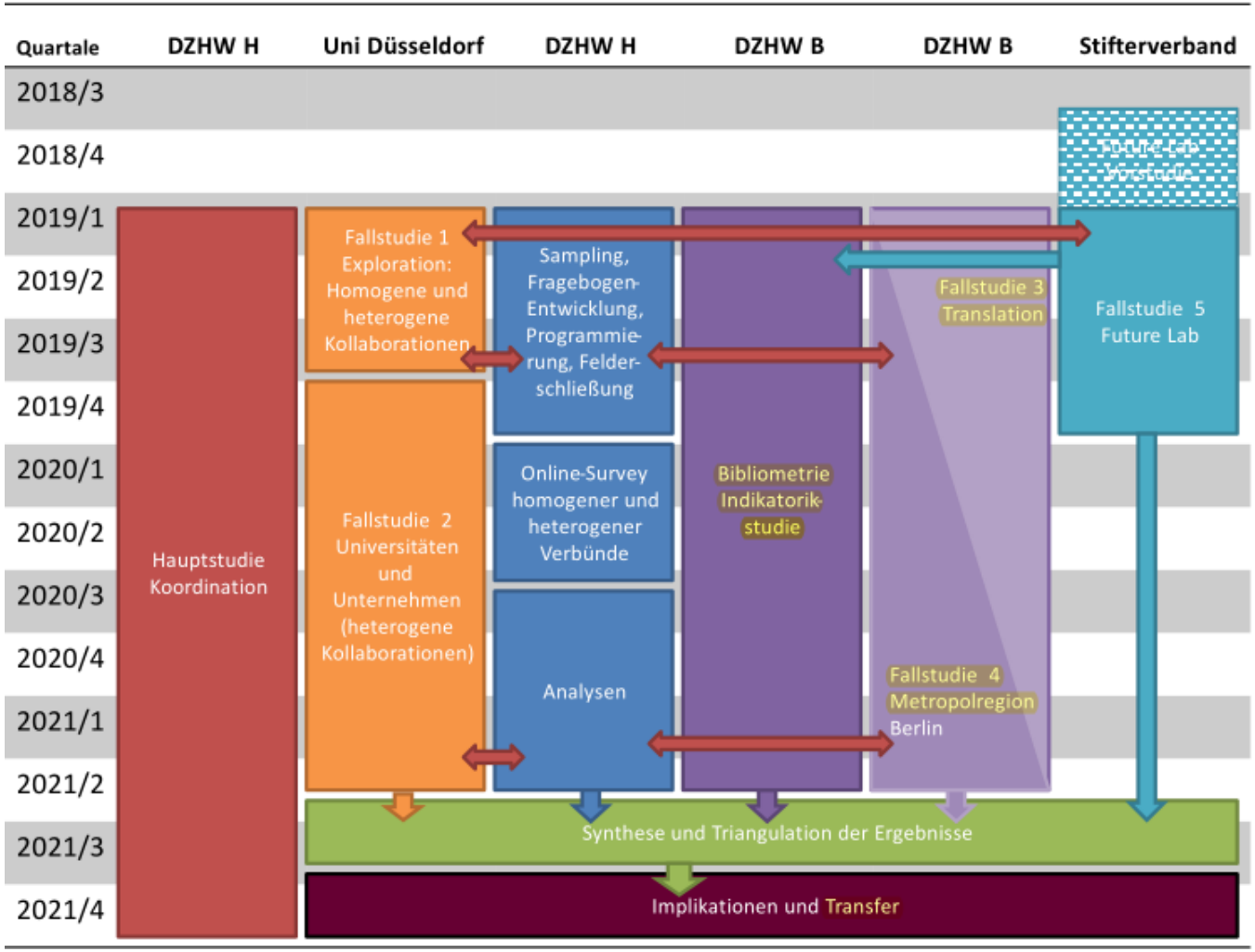
Outline of presentation

DZHW

1. Case study's guiding questions and DEKiF's bigger picture
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. Disciplinary view to scientific collaboration in Berlin region
5. A look at five main institutes in Berlin (*HU, FU, TU, CH, BIH*)
6. Bipartite community detection (case of Humanities)
7. Further questions to explore next

DEKiF case studies structure

- Today Berlin region case study, Bibliometrics case study on iDIV next.



Research questions

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration?
3. Does internationalization lead the way?
4. Or is it a matter of disciplines?
5. What spatial network analysis of coauthorships (organization level) would add?
6. Can we find more with Bipartite community detection of coauthorships networks?

Research questions

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration?
3. Does internationalization lead the way?
4. Or is it a matter of disciplines?
5. What spatial network analysis of coauthorships (organization level) would add?
6. Can we find more with Bipartite community detection of coauthorships networks?

Research questions

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration?
3. Does internationalization lead the way?
4. Or is it a matter of disciplines?
5. What spatial network analysis of coauthorships (organization level) would add?
6. Can we find more with Bipartite community detection of coauthorships networks?

Research questions

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration?
3. Does internationalization lead the way?
4. Or is it a matter of disciplines?
5. What spatial network analysis of coauthorships (organization level) would add?
6. Can we find more with Bipartite community detection of coauthorships networks?

Research questions

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration?
3. Does internationalization lead the way?
4. Or is it a matter of disciplines?
5. What spatial network analysis of coauthorships (organization level) would add?
6. Can we find more with Bipartite community detection of coauthorships networks?

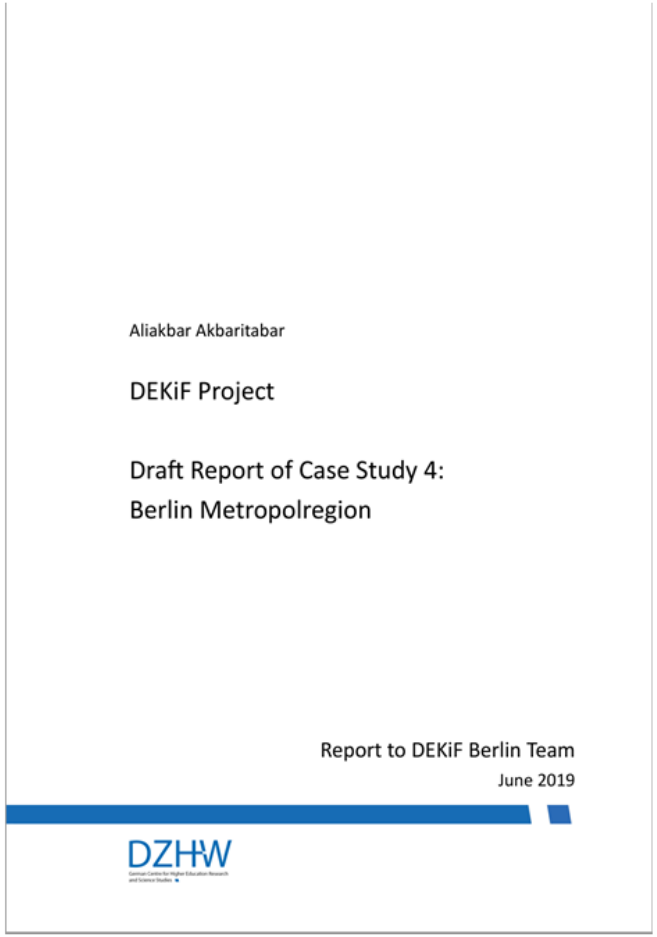
Research questions

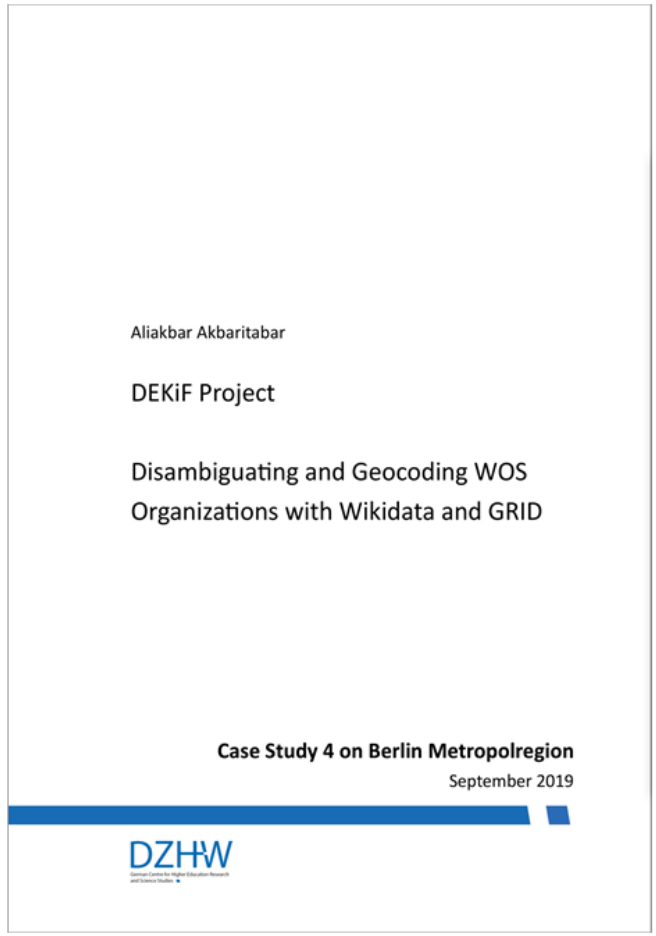
DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration?
3. Does internationalization lead the way?
4. Or is it a matter of disciplines?
5. What spatial network analysis of coauthorships (organization level) would add?
6. Can we find more with Bipartite community detection of coauthorships networks?

Two draft reports are prepared

- For DEKiF AP9-4 case study (previous preliminary phase)
- The report on matching has sample code to replicate





Brief on data and methods

DZHW

1. All WOS/Scopus **1990-2017** (b_2018 KB) pubs
2. *Article*, *Review* and *Conference proceeding* as **document types**
3. With at least one author/institution from **Berlin**, **Germany**
4. Number of publications, fractional count, 3 years citations, disciplines, journals, etc.
5. **Wikidata** 27th March 2019, **GRID** 17th February 2019 (10-12-2019 for address complements)
6. **ROR** organization registry and disambiguation API (12-2019)

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

Brief on data and methods

DZHW

1. All WOS/Scopus **1990-2017** (b_2018 KB) pubs
2. *Article*, *Review* and *Conference proceeding* as **document types**
3. With at least one author/institution from **Berlin**, **Germany**
4. Number of publications, fractional count, 3 years citations, disciplines, journals, etc.
5. **Wikidata** 27th March 2019, **GRID** 17th February 2019 (10-12-2019 for address complements)
6. **ROR** organization registry and disambiguation API (12-2019)

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

Brief on data and methods

DZHW

1. All WOS/Scopus **1990-2017** (b_2018 KB) pubs
2. *Article*, *Review* and *Conference proceeding* as **document types**
3. With at least one author/institution from **Berlin**, **Germany**
4. Number of publications, fractional count, 3 years citations, disciplines, journals, etc.
5. **Wikidata** 27th March 2019, **GRID** 17th February 2019 (10-12-2019 for address complements)
6. **ROR** organization registry and disambiguation API (12-2019)

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

Brief on data and methods

1. All WOS/Scopus **1990-2017** (b_2018 KB) pubs
2. *Article*, *Review* and *Conference proceeding* as **document types**
3. With at least one author/institution from **Berlin**, **Germany**
4. Number of publications, fractional count, 3 years citations, disciplines, journals, etc.
5. **Wikidata** 27th March 2019, **GRID** 17th February 2019 (10-12-2019 for address complements)
6. **ROR** organization registry and disambiguation API (12-2019)

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

Brief on data and methods

1. All WOS/Scopus **1990-2017** (b_2018 KB) pubs
2. *Article*, *Review* and *Conference proceeding* as **document types**
3. With at least one author/institution from **Berlin**, **Germany**
4. Number of publications, fractional count, 3 years citations, disciplines, journals, etc.
5. **Wikidata** 27th March 2019, **GRID** 17th February 2019 (10-12-2019 for address complements)
6. **ROR** organization registry and disambiguation API (12-2019)

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

Brief on data and methods

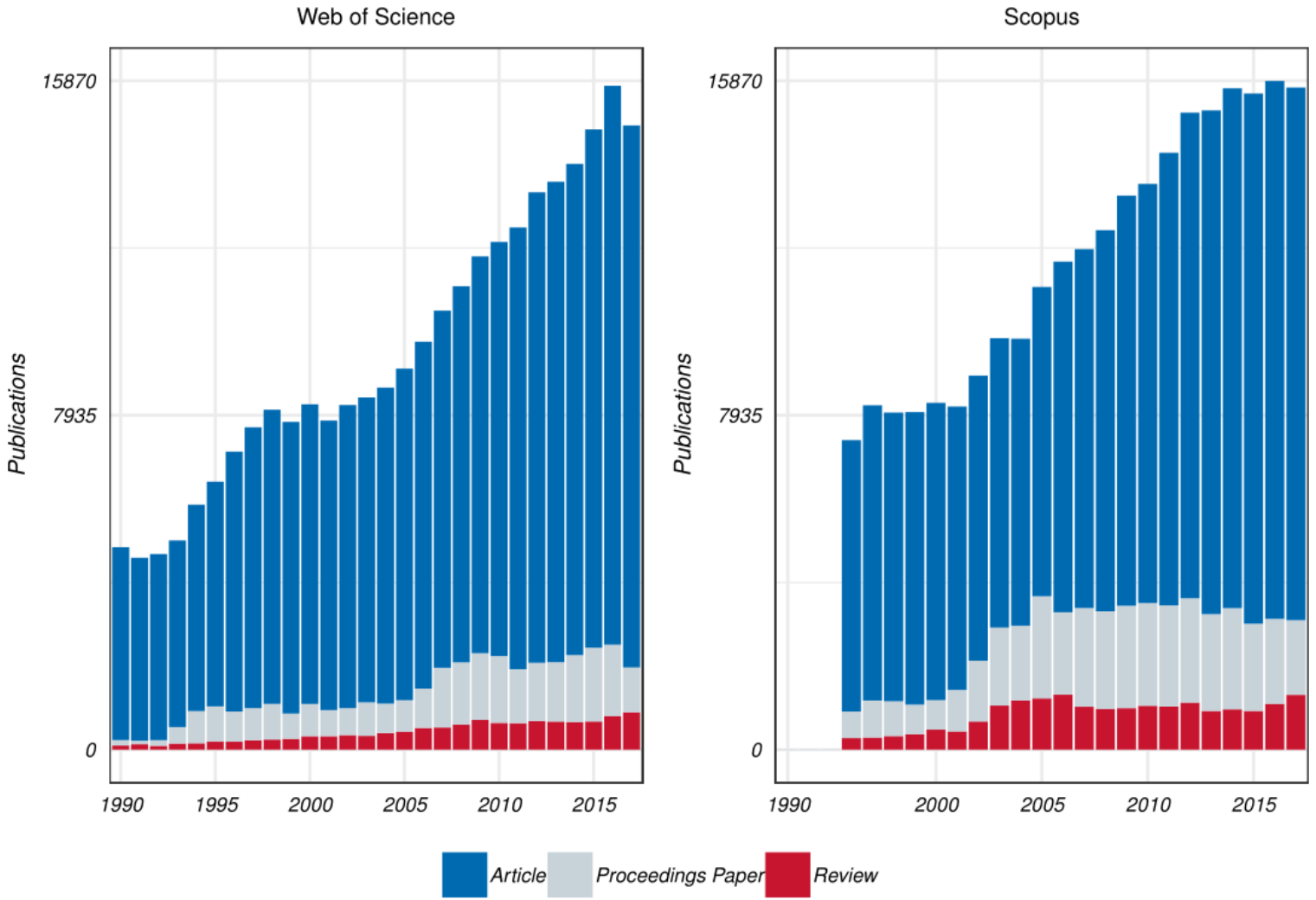
DZHW

1. All WOS/Scopus **1990-2017** (b_2018 KB) pubs
2. *Article*, *Review* and *Conference proceeding* as **document types**
3. With at least one author/institution from **Berlin**, **Germany**
4. Number of publications, fractional count, 3 years citations, disciplines, journals, etc.
5. **Wikidata** 27th March 2019, **GRID** 17th February 2019 (10-12-2019 for address complements)
6. **ROR** organization registry and disambiguation API (12-2019)

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

Berlin sample publications



German publications (baseline)

- Berlin follows/leads the general trend

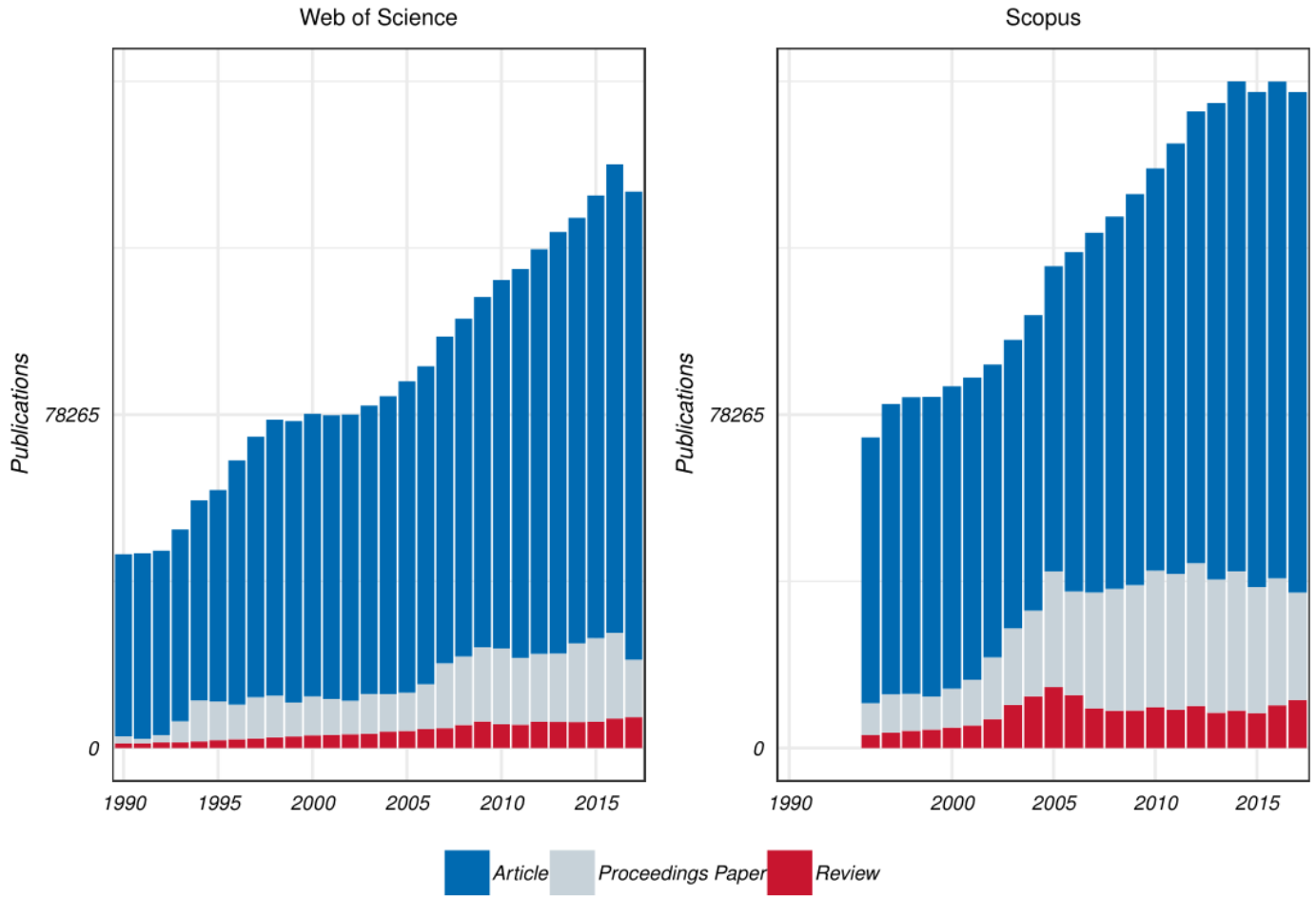


Figure 2: Articles, Reviews and Conference proceedings published by German institutes and universi-

In matching, Wikidata was limited

DZHW

- To instances of:
  - '*Comprehensive university*' (Q1767829)
  - '*Public university*' (Q875538)
  - '*University*' (Q3918)
  - '*Academic institution*' (Q4671277)
  - '*Fraunhofer Institute*' (Q20168706)
  - '*Research institute*' (Q31855)
  - '*Scientific society*' (Q748019)
  - '*Scientific organisation*' (Q45103187)
  - '*Max Planck Society*' (Q158085)
  - '*Max Planck Institute*' (Q6019423).
- These limited our data from over 55 million cases to **106,794** entities.
- In 2nd phase, all items with geogrphical coordinates (4,723,171 items) were used

Unique organizations (problematic?!)

- Organization name disambiguation, *why is it a must*?!

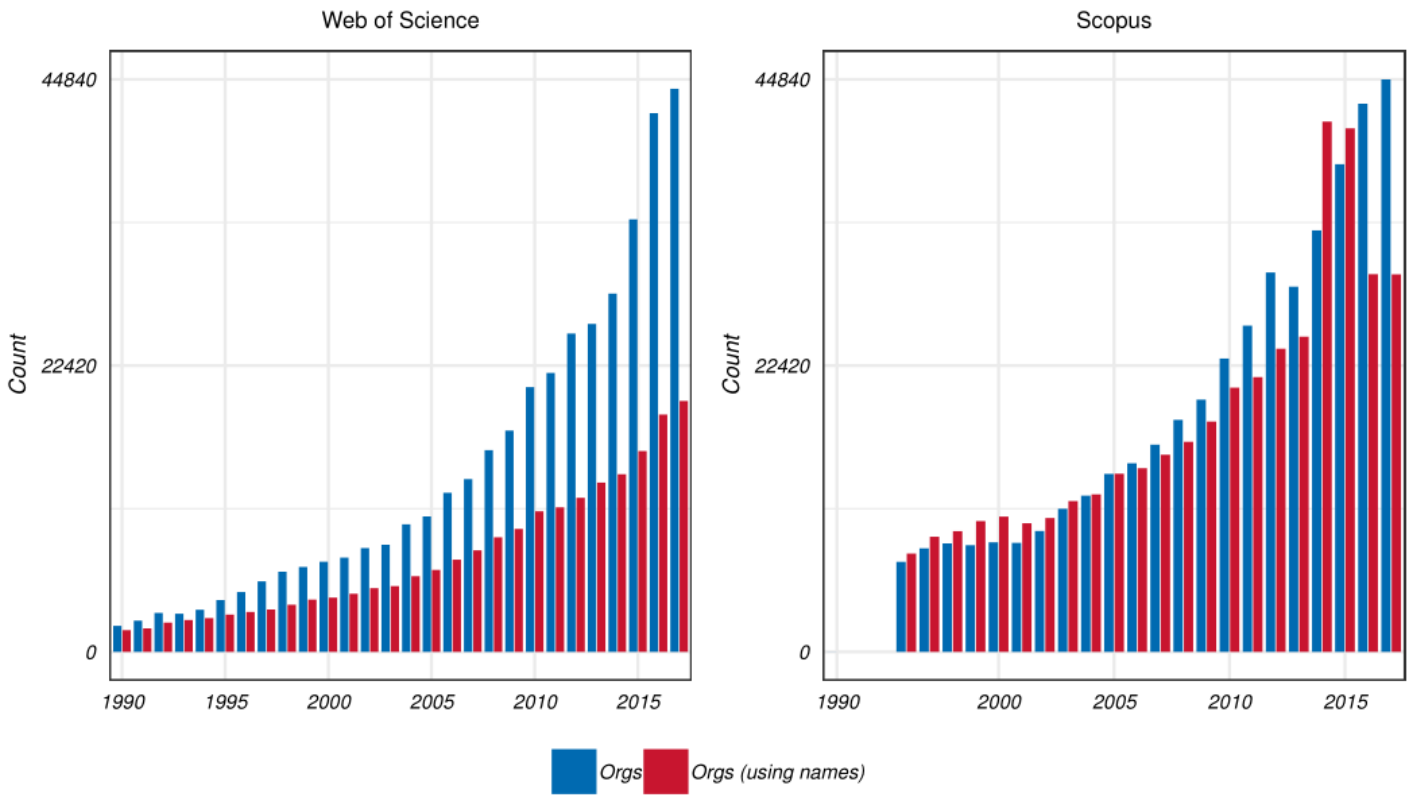


Figure 10: Unique organizations with which Berlin region institutes and universities have collaborated in Articles, Reviews and Conference proceedings in WOS and Scopus in 1990 - 2017

International organization example (1/2)

DZHW.

	FK_INSTITUTIONS	PK_KB_INST	KB_NAME	ORGANIZATION1
	All	All	All	aalborg univ
2	29009365	NA	NA	aalborg univ
3	1077601	NA	NA	aalborg univ
4	30422015	NA	NA	aalborg univ
6	8497964	NA	NA	aalborg univ
7	3823950	NA	NA	aalborg univ
8	5559997	NA	NA	aalborg univ
9	20140789	NA	NA	aalborg univ
10	2583771	NA	NA	aalborg univ
11	21284558	NA	NA	aalborg univ
12	27026789	NA	NA	aalborg univ
14	34366196	NA	NA	aalborg univ
15	23317559	NA	NA	aalborg univ
16	24242911	NA	NA	aalborg univ
17	28629165	NA	NA	aalborg univ
18	4170458	NA	NA	aalborg univ
19	6898547	NA	NA	aalborg univ
20	33059701	NA	NA	aalborg univ
22	24567558	NA	NA	aalborg univ
23	18351477	NA	NA	aalborg univ
24	17626795	NA	NA	aalborg univ
25	22866279	NA	NA	aalborg univ
26	7812723	NA	NA	aalborg univ
27	18269492	NA	NA	aalborg univ
28	14067026	NA	NA	aalborg univ
30	27769743	NA	NA	aalborg univ
31	18178514	NA	NA	aalborg univ
33	4883551	NA	NA	aalborg univ
34	16148970	NA	NA	aalborg univ
35	26011466	NA	NA	aalborg univ
36	23019925	NA	NA	aalborg univ
38	18861657	NA	NA	aalborg univ
39	20241696	NA	NA	aalborg univ

fuzzy_match_wiki_id	fuzzy_match_wiki_name	fuzzy_jw_level
All	All	All
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1
Q601956	aalborg university	0.85 < JW < 1

German organization example (2/2)

DZHW

	FK_INSTITUTIONS	PK_KB_INST	KB_NAME	ORGANIZATION1	CITY	COUNTRYCODE	POSTALCODE
1	24966247	NA	NA	alexander von humboldt inst internet & gesell	berlin	deu	D-10117
2	26263851	NA	NA	alexander von humboldt inst internet & gesell	berlin	deu	NA
3	25284785	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
4	19041909	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	D-10117
5	23459193	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
6	9814790	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	D-10117
7	5548014	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
8	32465471	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	D-10117
9	6357212	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
10	2595255	NA	NA	alexander von humboldt inst internet & soc hiig	berlin	deu	NA

fuzzy_match_wiki_id	fuzzy_match_wiki_name	fuzzy_jw_level	fuzzy_city_status
NA	NA	NA	NA
NA	NA	NA	NA
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches

Only matched with English Wikipedia, "gesell"

Encoding (problematic?!)

DZHW

Worksheet | Query Builder

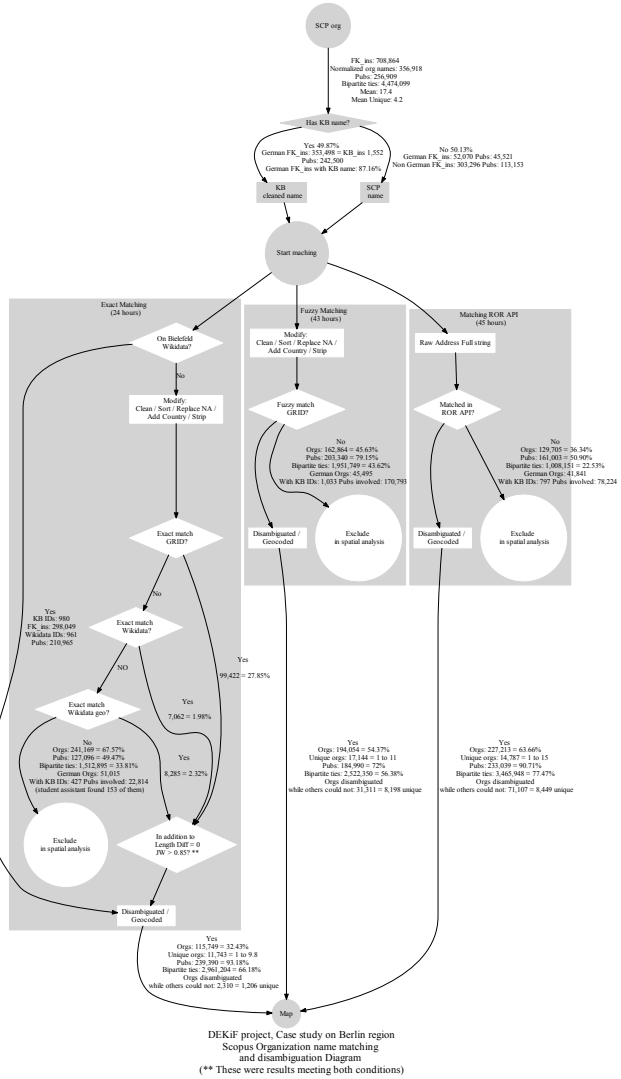
```
select PK_INSTITUTIONS, ORGANIZATION1, INSTITUTION_FULL from scopus_b_2018.institutions
where pk_institutions in ('10897353', '29324169', '14472460', '46335179', '31666249', '696390', '40120836','27118260', '28089912',
```

Query Result

SQL | All Rows Fetched: 10 in 0.004 seconds

	PK_INSTITUTIONS	ORGANIZATION1	INSTITUTION_FULL
1	696390	Thï¿½ringer Landessternwarte Tautenburg	Thï¿½ringer Landessternwarte Tautenburg
2	6942610	Universität Bern	Universität Bern, Institute of Psychology
3	10897353	Charitï¿½ - Universitï¿½tsmedizin	Charitï¿½ - Universitï¿½tsmedizin, Department of Neurology
4	14472460	UniversitätSiegen	UniversitätSiegen, Fachbereich Physik
5	27118260	Bundestierärztekammer e.V	Bundestierärztekammer e.V
6	28089912	RÖNTEC GmbH	RÖNTEC GmbH
7	29324169	Max-Delbrï¿½ck-Center for Molecular Medicine	Max-Delbrï¿½ck-Center for Molecular Medicine
8	31666249	University of Tï¿½bingen	University of Tï¿½bingen, Institute of Tropical Medicine
9	40120836	Else-Kröner-Fresenius-Zentrum	Else-Kröner-Fresenius-Zentrum
10	46335179	Institut für Ärztliche Begutachtung	Institut für Ärztliche Begutachtung

Disambiguation & geocoding logic & results



Comparison of different disambiguations

DZHW

X	Non_disamb	Exact	Fuzzy	ROR
Number of connected components	10,269	181	696	282
Number of biparitite nodes	613,827	251,133	202,134	247,826
Number of biparitite edges	1,083,775	670,309	561,431	751,833
Number of biparitite nodes G	582,958	250,715	200,498	247,168
Number of biparitite nodes G %	95	100	99	100
Number of biparitite edges G	1,063,001	670,071	560,487	751,455
Number of biparitite edges G %	98	100	100	100
Density G	0	0	0	0
Number orgs	356,918	11,743	17,144	14,787
Number orgs G	337,755	11,551	16,419	14,484
Number papers	256,909	239,390	184,990	233,039
Number papers G	245,203	239,164	184,079	232,684

Some descriptive figures

DZHW

- If pubs $>$ 20,000 **name** on map, if 1,000 $<$ pubs $<$ 20,000 **number** on map
- KB Sectors (left) vs. GRID/ROR organization types (right)

index	SECTOR
nan	14223
Sonstige	176
Hochschulen	131
Max-Planck-Gesellschaft	62
Wirtschaft	59
Leibniz-Gemeinschaft	53
Fraunhofer-Gesellschaft	41
Ressortforschung	32
Helmholtz-Gemeinschaft	10

index	org_type
Education	4281
Healthcare	2980
Facility	2960
Company	1600
Nonprofit	1105
Government	899
Other	634
Archive	221
nan	107

Organization types (contrasting both)

DZHW

GRID.org.type	KB.Sector	Count_excluding_NA
Archive	Leibniz-Gemeinschaft	5
Archive	Max-Planck-Gesellschaft	1
Archive	Ressortforschung	2
Archive	Sonstige	4
Archive	Wirtschaft	1
Company	Hochschulen	1
Company	Sonstige	3
Company	Wirtschaft	38
Education	Hochschulen	122
Education	Leibniz-Gemeinschaft	1
Education	Max-Planck-Gesellschaft	1
Education	Sonstige	6
Education	Wirtschaft	3
Facility	Fraunhofer-Gesellschaft	39
Facility	Helmholtz-Gemeinschaft	9
Facility	Leibniz-Gemeinschaft	43
Facility	Max-Planck-Gesellschaft	58
Facility	Ressortforschung	12
Facility	Sonstige	32
Facility	Wirtschaft	10
Government	Fraunhofer-Gesellschaft	1
Government	Helmholtz-Gemeinschaft	1
Government	Ressortforschung	15
Government	Sonstige	10
Healthcare	Hochschulen	8
Healthcare	Ressortforschung	2
Healthcare	Sonstige	92
Healthcare	Wirtschaft	1
Nonprofit	Fraunhofer-Gesellschaft	1
Nonprofit	Leibniz-Gemeinschaft	2
Nonprofit	Max-Planck-Gesellschaft	2
Nonprofit	Ressortforschung	1
Nonprofit	Sonstige	18
Nonprofit	Wirtschaft	4
Other	Leibniz-Gemeinschaft	2
Other	Sonstige	11
Other	Wirtschaft	2

List of Disciplines (and abbreviations)

DZHW

- From **OECD** mapping by Stephan using **Scopus ASJC** categories
- Some publications are assigned to **multiple disciplines**, that is why on maps they appear multiple times (as interdisciplinary collaboration)
- *Please remember these abbreviations*:
  1. Agricultural Sciences = '**AS**'
  2. Engineering Technology = '**ET**'
  3. Humanities = '**H**'
  4. Medical Health Sciences = '**MHS**'
  5. Natural Sciences = '**NS**'
  6. Social Sciences = '**SS**'

Comparison of Disciplines

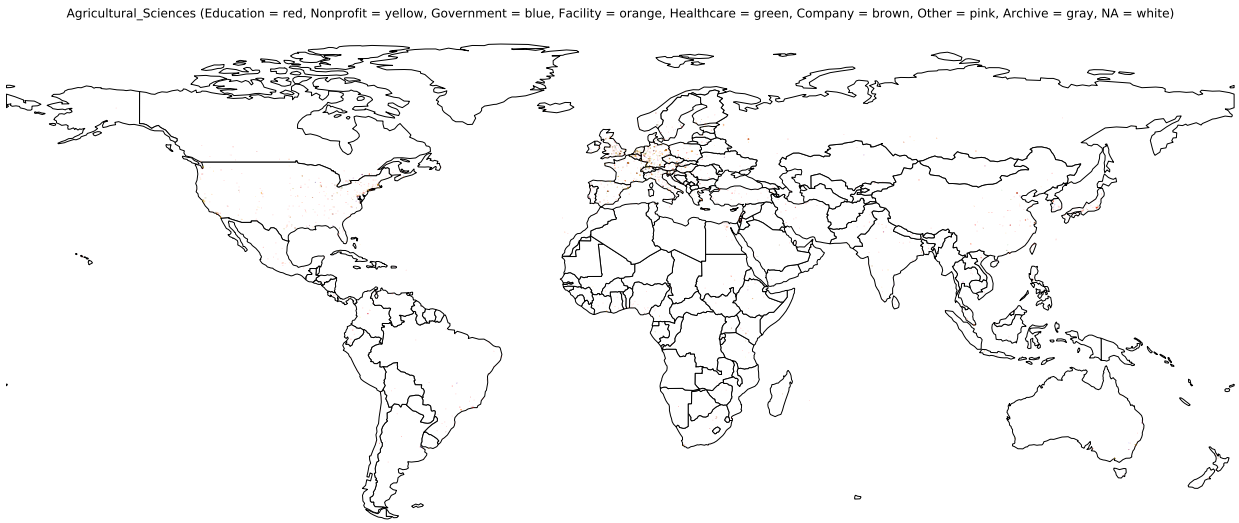
DZHW

X	AS	ET	H	MHS	NS	SS
Number of connected components	80	195	77	126	208	165
Number of biparitite nodes	17,130	59,859	6,376	99,549	164,105	27,363
Number of biparitite edges	32,971	144,680	8,791	263,586	542,856	54,628
Number of biparitite nodes G	16,939	59,387	6,197	99,262	163,631	26,966
Number of biparitite nodes G %	99	99	97	100	100	99
Number of biparitite edges G	32,858	144,402	8,688	263,424	542,589	54,395
Number of biparitite edges G %	100	100	99	100	100	100
Density G	0	0	0	0	0	0
Number orgs	3,678	5,190	1,165	9,403	11,541	4,378
Number orgs G	3,585	4,979	1,078	9,263	11,321	4,191
Number papers	13,452	54,669	5,211	90,146	152,564	22,985
Number papers G	13,354	54,408	5,119	89,999	152,310	22,775

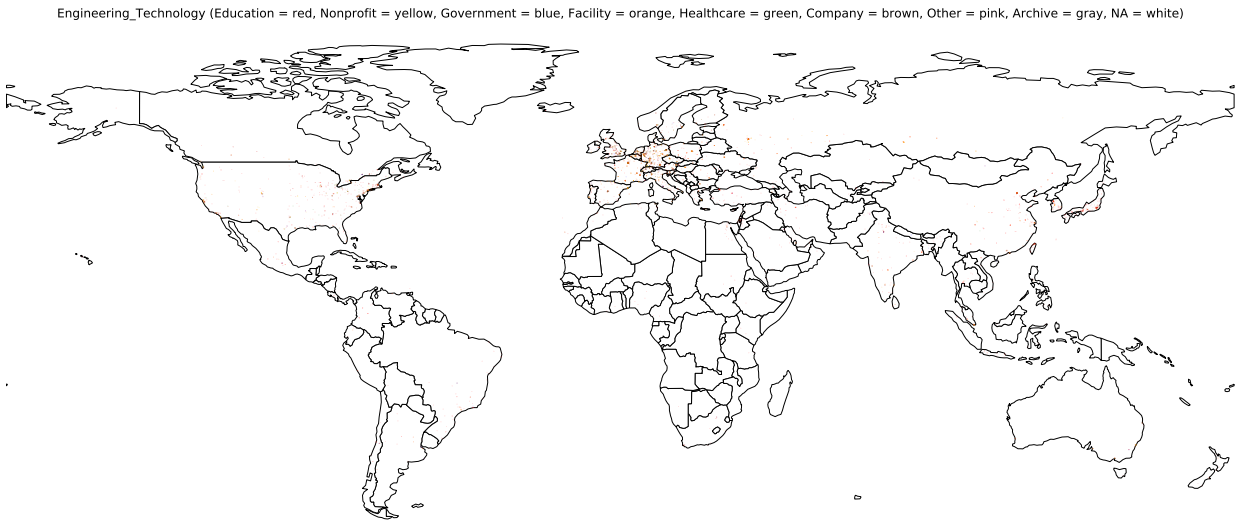
A legend of organization colors

- Education = red
- Nonprofit = yellow
- Government = blue
- Facility = orange
- Healthcare = green
- Company = brown
- Other = pink
- Archive = gray
- NA = white

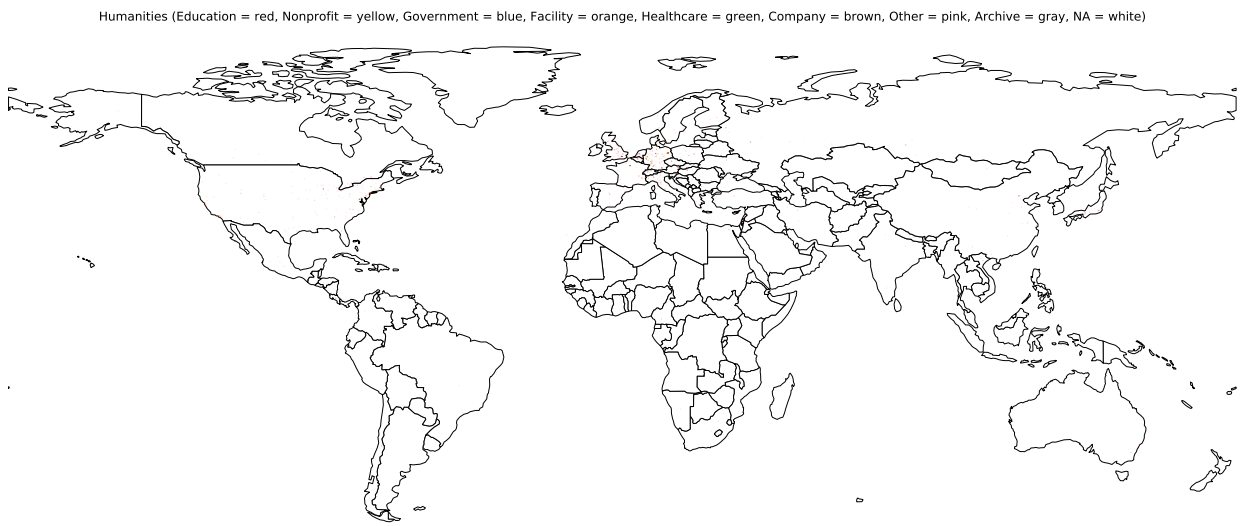
Map AS


Agricultural_Sciences (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

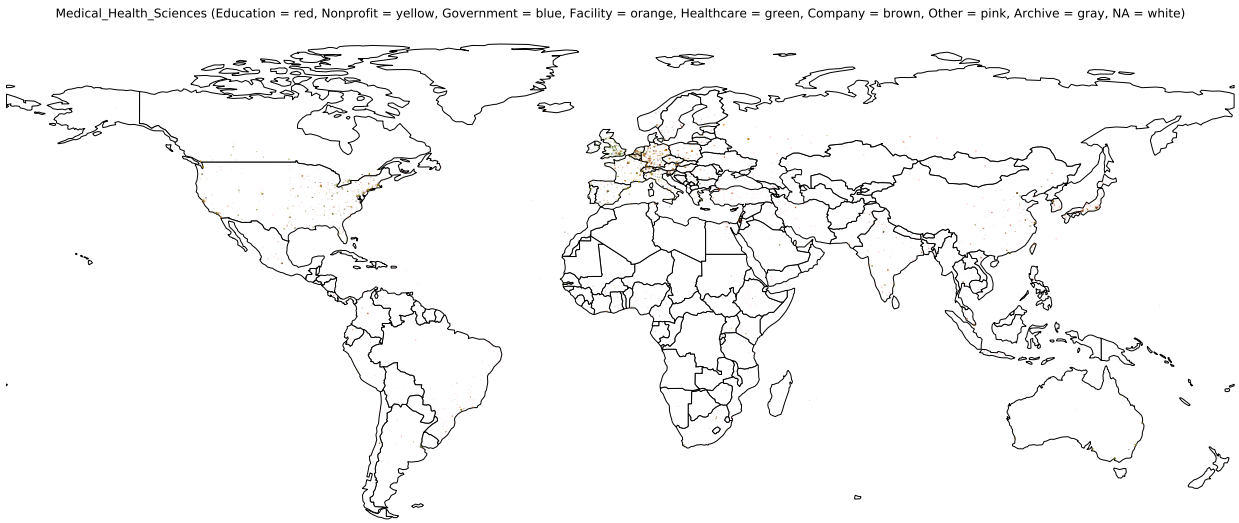
Map ET


Engineering_Technology (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

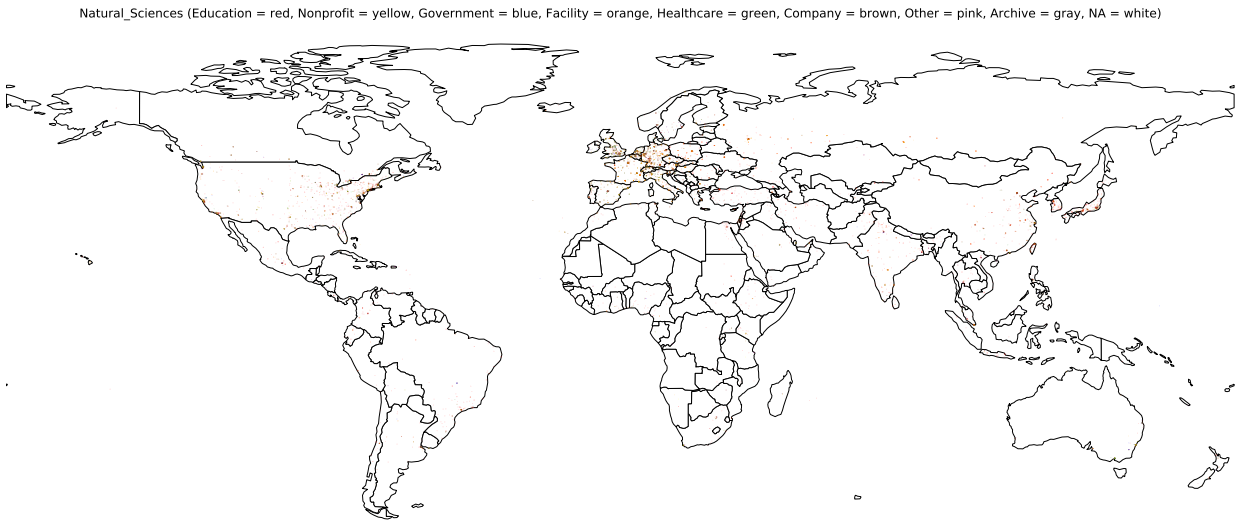
Map H


Humanities (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

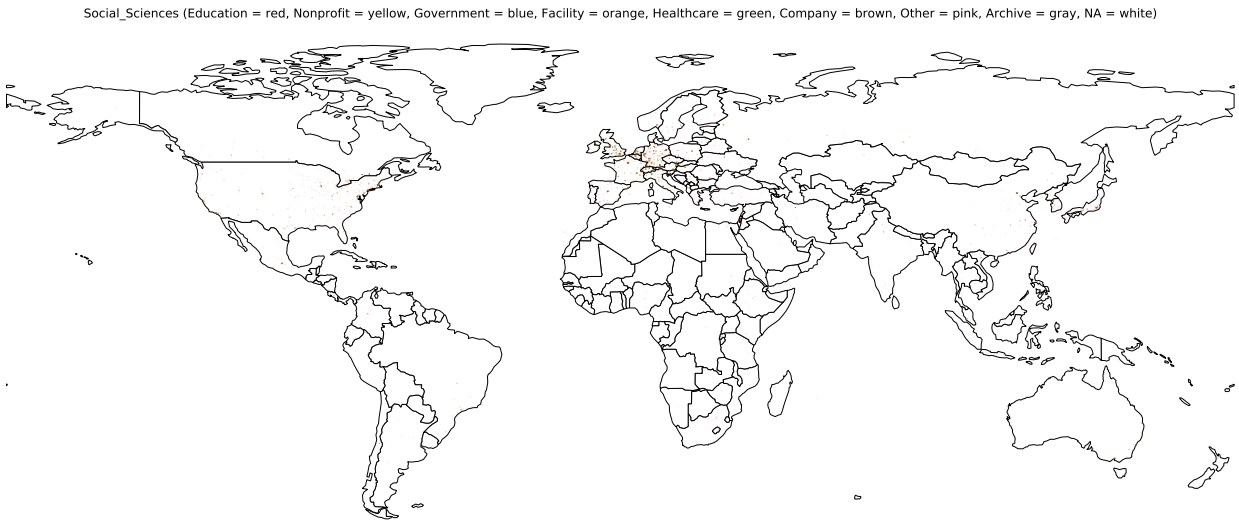
Map MHS


Medical_Health_Sciences (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

Map NS


Natural_Sciences (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

Map SS


Social_Sciences (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

Five main institutes in Berlin

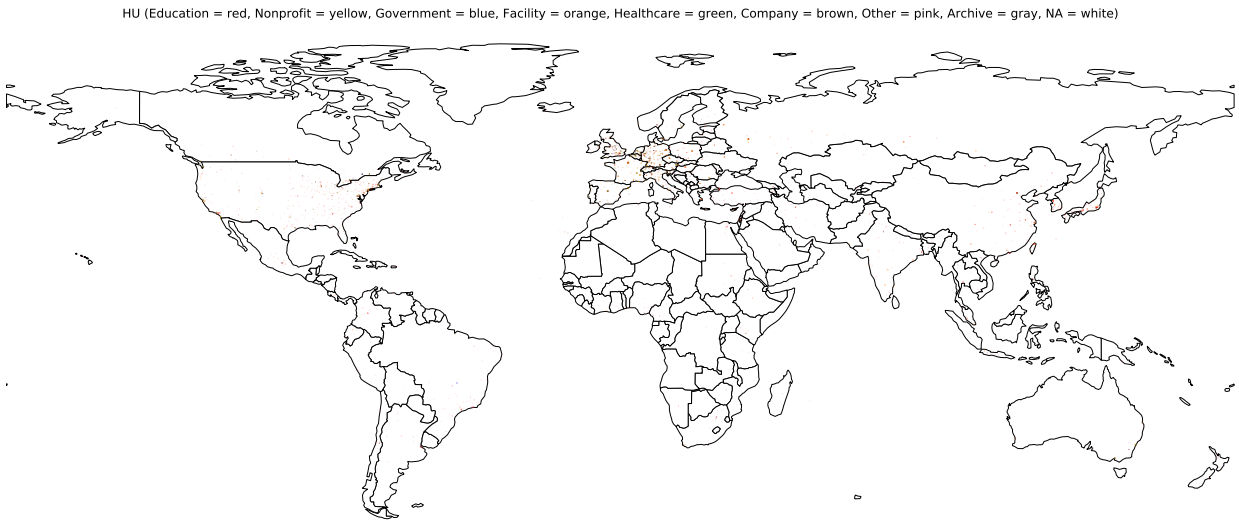
- HU_berlin = 'HU'
- FU_berlin = 'FU'
- TU_berlin = 'TU'
- CH_berlin (charite) = 'CH'
- BIH_berlin = 'BIH'

Five Berlin institutes networks

DZHW

X	HU	FU	TU	CH	BIH	Union.5
Number of connected components	1	1	1	1	1	1
Number of biparitite nodes	39,402	37,825	37,669	66,667	976	161,253
Number of biparitite edges	252,426	80,688	71,220	189,332	2,157	553,641
Number of biparitite nodes G	39,402	37,825	37,669	66,667	976	161,253
Number of biparitite nodes G %	100	100	100	100	100	100
Number of biparitite edges G	252,426	80,688	71,220	189,332	2,157	553,641
Number of biparitite edges G %	100	100	100	100	100	100
Density G	0	0	0	0	0	0
Number orgs	4,913	4,514	3,655	6,955	622	11,364
Number orgs G	4,913	4,514	3,655	6,955	622	11,364
Number papers	34,489	33,311	34,014	59,712	354	149,889
Number papers G	34,489	33,311	34,014	59,712	354	149,889

Map HU


HU (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

Map FU


FU (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

Map TU


TU (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

Map CH


CH (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

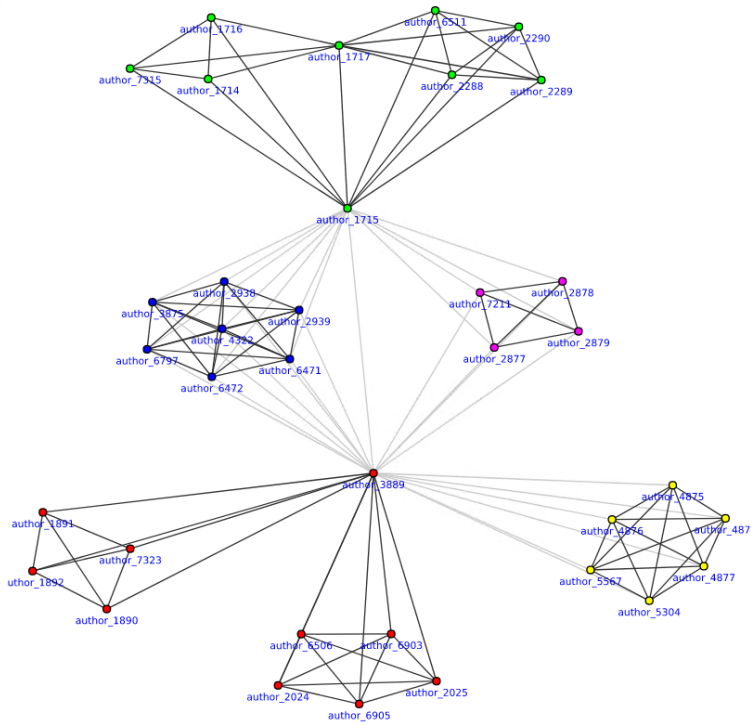
Map BIH


BIH (Education = red, Nonprofit = yellow, Government = blue, Facility = orange, Healthcare = green, Company = brown, Other = pink, Archive = gray, NA = white)

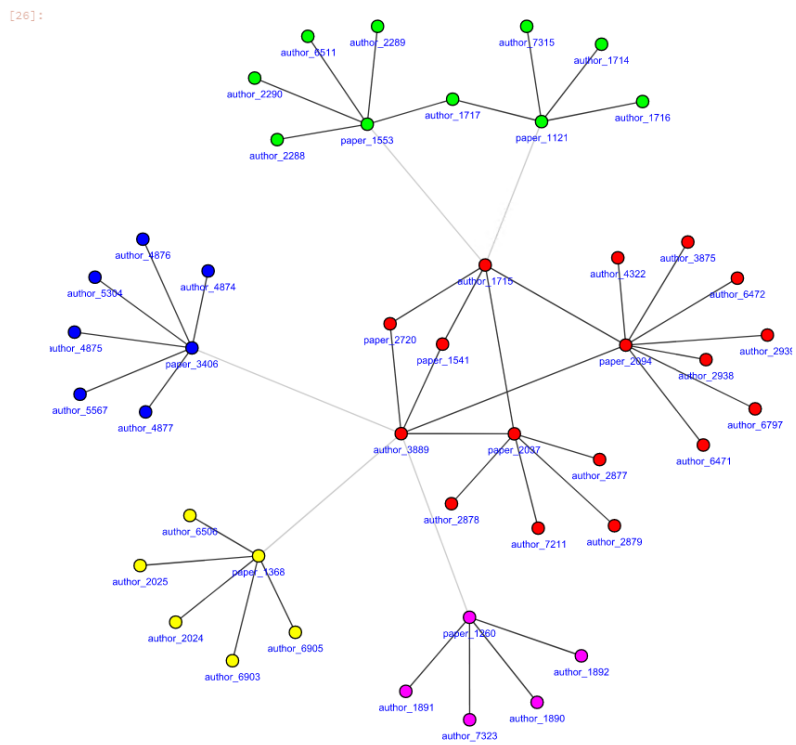
Bipartite community detection (1/2)

DZHW

[123]:



[26]:



[27]: `ig.plot(g1)`

Bipartite community detection (2/2)

DZHW

- Constant Pots Model in `Leidenalg` library
- $\gamma =$ `5e-5`
- Yield 161 communities (here only first 6)

Cluster	Count.Org.Pubs
0	2,189
1	1,486
2	600
3	509
4	310
5	278

type2	cluster	COUNTRYCODE	unique_id
org	0	DEU	59
org	1	DEU	45
org	2	DEU	24
org	3	DEU	15
org	4	DEU	4
org	5	DEU	12

Communities of network structure (Humanities)

DZHW

- One specific case, cluster 1 in Humanities G comp.

Cluster	Organization.or.Paper	Count.unique
1	org	237
1	paper	1249

ASJC.DESCRIPTION	Count.Pubs
Archeology	112
Archeology (arts and humanities)	148
Arts and Humanities (miscellaneous)	168
Classics	9
Conservation	2
General Arts and Humanities	16
History	234
History and Philosophy of Science	55
Language and Linguistics	90
Linguistics and Language	121
Literature and Literary Theory	77
Music	34
Philosophy	109
Religious Studies	32
Visual Arts and Performing Arts	42

More on H Cluster 1

Total.pubs.number	Count.orgs
1	113
2	36
3	24
4	11
5	8
6	8
7	11
8	3
9	2
10	1
11	1
12	1
13	2
14	1
15	3
17	1
20	1
22	1
24	3
27	1
28	2
36	1
161	1
1,235	1

Citations.number.in.first.3.years	Count.Pubs
0	878
1	155
2	66
3	43
4	25
5	23
6	8
7	13
8	10
9	8
10	5
11	6
12	1
15	2
18	1
20	1
22	1
24	1
29	1
68	1

More on H Cluster 1

COUNTRYCODE	Count.orgs
ARG	2
AUS	4
AUT	3
AZE	1
BGR	1
BRA	2
CAN	4
CHE	5
CHN	11
CRI	2
CZE	1
DEU	45
DNK	2
EGY	2
ESP	4
EST	1
FRA	5
GBR	28
GRC	3
HRV	1
IND	3
IRN	3
ISR	4
ITA	7
JPN	11
KOR	1
LKA	1
LUX	1

Take home messages

DZHW

1. Need for lengthy & time consuming disambiguation
2. It is a must as 1 in 8 WOS (1/10 or 1/11 or 1/15 SCP) unique organization IDs proved reliable
3. Network analysis view to collaboration, composition & temporal evolution will be biased without disambiguation

Take home messages

DZHW

1. Need for lengthy & time consuming disambiguation
2. It is a must as 1 in 8 WOS (1/10 or 1/11 or 1/15 SCP) unique organization IDs proved reliable
3. Network analysis view to collaboration, composition & temporal evolution will be biased without disambiguation

Take home messages

DZHW

1. Need for lengthy & time consuming disambiguation
2. It is a must as 1 in 8 WOS (1/10 or 1/11 or 1/15 SCP) unique organization IDs proved reliable
3. Network analysis view to collaboration, composition & temporal evolution will be biased without disambiguation

Answering research questions (preliminary)

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration? **seems so**
3. Does internationalization lead the way? **for some countries**
4. Or is it a matter of disciplines? **to a high extent**
5. What spatial network analysis of coauthorships (organization level) would add? **I think so**
6. Can we find more with Bipartite community detection of coauthorships networks? **Be the judge**

Answering research questions (preliminary)

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration? **seems so**
3. Does internationalization lead the way? **for some countries**
4. Or is it a matter of disciplines? **to a high extent**
5. What spatial network analysis of coauthorships (organization level) would add? **I think so**
6. Can we find more with Bipartite community detection of coauthorships networks? **Be the judge**

Answering research questions (preliminary)

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration? **seems so**
3. Does internationalization lead the way? **for some countries**
4. Or is it a matter of disciplines? **to a high extent**
5. What spatial network analysis of coauthorships (organization level) would add? **I think so**
6. Can we find more with Bipartite community detection of coauthorships networks? **Be the judge**

Answering research questions (preliminary)

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration? **seems so**
3. Does internationalization lead the way? **for some countries**
4. Or is it a matter of disciplines? **to a high extent**
5. What spatial network analysis of coauthorships (organization level) would add? **I think so**
6. Can we find more with Bipartite community detection of coauthorships networks? **Be the judge**

Answering research questions (preliminary)

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration? **seems so**
3. Does internationalization lead the way? **for some countries**
4. Or is it a matter of disciplines? **to a high extent**
5. What spatial network analysis of coauthorships (organization level) would add? **I think so**
6. Can we find more with Bipartite community detection of coauthorships networks? **Be the judge**

Answering research questions (preliminary)

DZHW

1. What is *Homogeneous* or *Heterogeneous* collaboration here?
2. Is there a sectoral division in collaboration? **seems so**
3. Does internationalization lead the way? **for some countries**
4. Or is it a matter of disciplines? **to a high extent**
5. What spatial network analysis of coauthorships (organization level) would add? **I think so**
6. Can we find more with Bipartite community detection of coauthorships networks? **Be the judge**

More questions to explore

- **This is a work in progress**, please suggest further questions to explore
- My further questions:
- If there are policies guiding scientific collaboration causing current sector and discipline based collaboration (which needs further validation)
- Who are the international organizations chosen by Berlin institutes for collaboration? Are they the high permorers in each country? (e.g., Patient exchange by hospitals)
- Further analyze the homophily effects (organizations activity, academic age, first and last publications, spatial/geographical proximity, country, continent, etc.)
- Which specific cases (organization, country, alliance) could be explored further? (e.g., see Turkey? Italy?)
- How to explore university and organization profiles and priorities through this data, is homophily evaluation enough?
- Funding information of these publications needs to be added still.
- What else?
- Wanna collaborate? Send me your Gitlab handle.

Thanks for your attention!

When filled ROR network with non-disambiguated nodes

	Non_disamb	Exact	Fuzzy	ROR	ROR_Filled
Clusters_num	10,269	181	696	373	6,735
Vcount	613,827	251,133	202,134	227,556	401,402
Ecount	1,083,775	670,309	561,431	694,322	1,042,008
G_Vcount	582,958	250,715	200,498	226,683	384,213
G_Vcount_Percent	94.97%	99.83%	99.19%	99.62%	95.72%
G_Ecount	1,063,001	670,071	560,487	693,820	1,031,511
G_Ecount_Percent	98.08%	99.96%	99.83%	99.93%	98.99%
G_Density	0.00001	0.00002	0.00003	0.00003	0.00001