Homogeneous and heterogeneous scientific collaboration

(DEKiF case study 4 and bibliometrics)

Aliakbar Akbaritabar (Ali)

Akbaritabar@DZHW.eu

27-28 February, 2020, Dusseldorf

DZHW
German Centre for Higher Education Research and Science Studies

Outline of presentation

DZHW

1. Case study 4 on Berlin region
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. A look at Heterogeneity of collaboration
5. Bipartite community detection
6. Bibliometric case study on iDIV (an introduction)

Outline of presentation

DZHW

1. Case study 4 on Berlin region
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. A look at Heterogeneity of collaboration
5. Bipartite community detection
6. Bibliometric case study on iDIV (an introduction)

Outline of presentation

DZHW

1. Case study 4 on Berlin region
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. A look at Heterogeneity of collaboration
5. Bipartite community detection
6. Bibliometric case study on iDIV (an introduction)

Outline of presentation

DZHW

1. Case study 4 on Berlin region
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. A look at Heterogeneity of collaboration
5. Bipartite community detection
6. Bibliometric case study on iDIV (an introduction)

Outline of presentation

DZHW

1. Case study 4 on Berlin region
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. A look at Heterogeneity of collaboration
5. Bipartite community detection
6. Bibliometric case study on iDIV (an introduction)

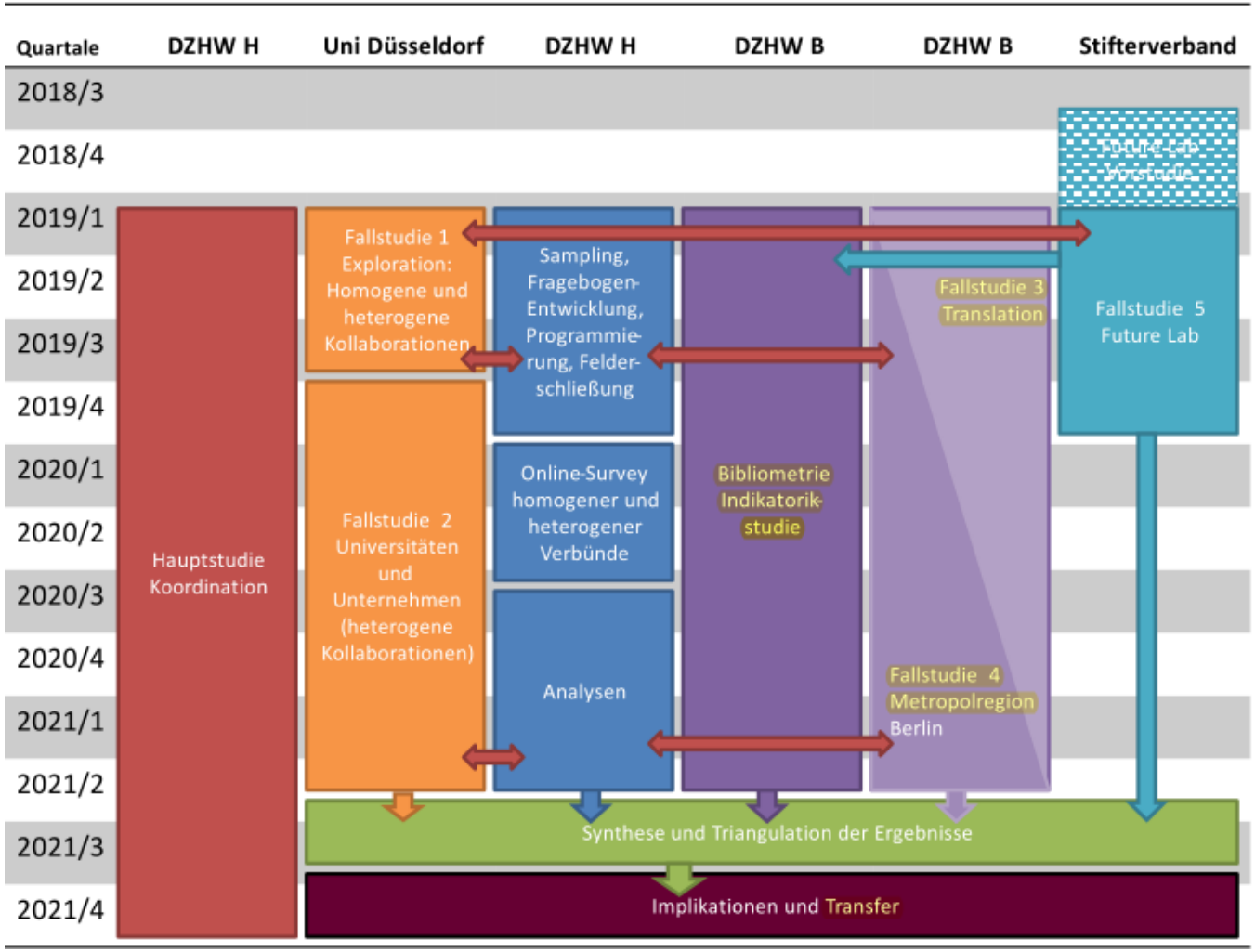
Outline of presentation

DZHW

1. Case study 4 on Berlin region
2. Data and methods
3. Organization name disambiguation, *why is it a must*?!
4. A look at Heterogeneity of collaboration
5. Bipartite community detection
6. Bibliometric case study on iDIV (an introduction)

DEKiF case studies structure

- Berlin region case study (analysis still in progress)
- A brief view to Bibliometric case study on iDIV



Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Research questions

DZHW

1. What forms can **scientific collaboration** take? (e.g., co-authorship, co-invention, co-funded, . . .)
2. What is *Homogeneous* or *Heterogeneous* collaboration?
3. *Heterogeneity* based on organization sectors?
4. *Heterogeneity* based on internationalization?
5. *Heterogeneity* based on disciplines?
6. *Heterogeneity* based on spatial proximity?
7. *Heterogeneity* based on clusters of co-authorships network?

Organizational scenarios of collaboration

DZHW

1. **Network** based (e.g., I. UNIBUND (ancestor of iDIV from 1995), II. iDIV as a fuzzy network)
2. **Organization** based (e.g., I. iDIV as a DFG founded center. II. BIH at highest level)
3. **Project** based (e.g., CRG/TRG projects in BIH)
4. **Region** based (e.g., Berlin metro-pole region)

Organizational scenarios of collaboration

DZHW

1. **Network** based (e.g., I. UNIBUND (ancestor of iDIV from 1995), II. iDIV as a fuzzy network)
2. **Organization** based (e.g., I. iDIV as a DFG founded center. II. BIH at highest level)
3. **Project** based (e.g., CRG/TRG projects in BIH)
4. **Region** based (e.g., Berlin metro-pole region)

Organizational scenarios of collaboration

DZHW

1. **Network** based (e.g., I. UNIBUND (ancestor of iDIV from 1995), II. iDIV as a fuzzy network)
2. **Organization** based (e.g., I. iDIV as a DFG founded center. II. BIH at highest level)
3. **Project** based (e.g., CRG/TRG projects in BIH)
4. **Region** based (e.g., Berlin metro-pole region)

Organizational scenarios of collaboration

DZHW

1. **Network** based (e.g., I. UNIBUND (ancestor of iDIV from 1995), II. iDIV as a fuzzy network)
2. **Organization** based (e.g., I. iDIV as a DFG founded center. II. BIH at highest level)
3. **Project** based (e.g., CRG/TRG projects in BIH)
4. **Region** based (e.g., Berlin metro-pole region)

Brief on data and methods

1. All 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** (Global Research Identifier Database) 17th February 2019 (10th December 2019 for address complements)
6. **Research Organization Registry (ROR)** [local] API (18th December 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 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** (Global Research Identifier Database) 17th February 2019 (10th December 2019 for address complements)
6. **Research Organization Registry (ROR)** [local] API (18th December 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 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** (Global Research Identifier Database) 17th February 2019 (10th December 2019 for address complements)
6. **Research Organization Registry (ROR)** [local] API (18th December 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 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** (Global Research Identifier Database) 17th February 2019 (10th December 2019 for address complements)
6. **Research Organization Registry (ROR)** [local] API (18th December 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 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** (Global Research Identifier Database) 17th February 2019 (10th December 2019 for address complements)
6. **Research Organization Registry (ROR)** [local] API (18th December 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 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** (Global Research Identifier Database) 17th February 2019 (10th December 2019 for address complements)
6. **Research Organization Registry (ROR)** [local] API (18th December 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

German publications (baseline)

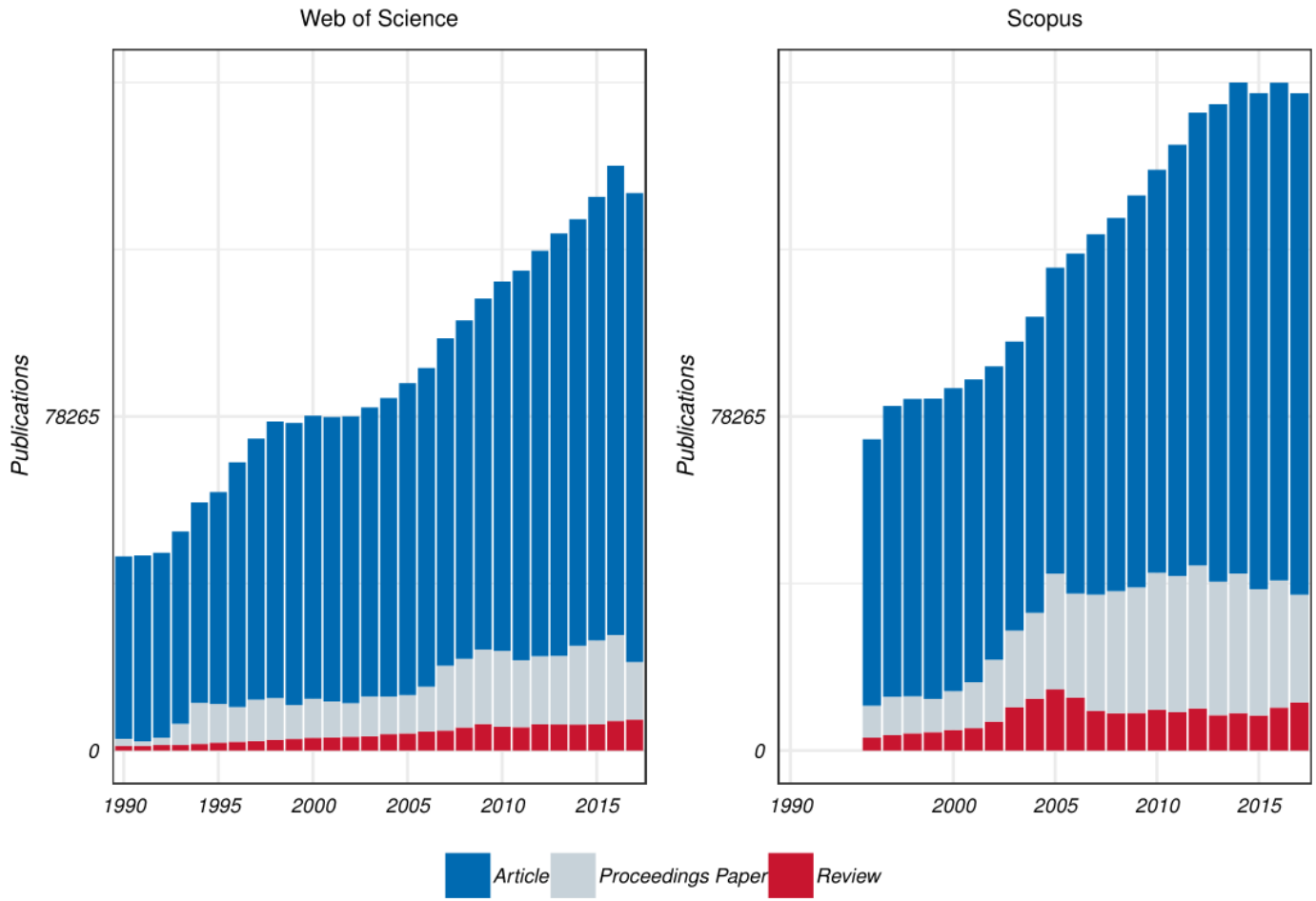
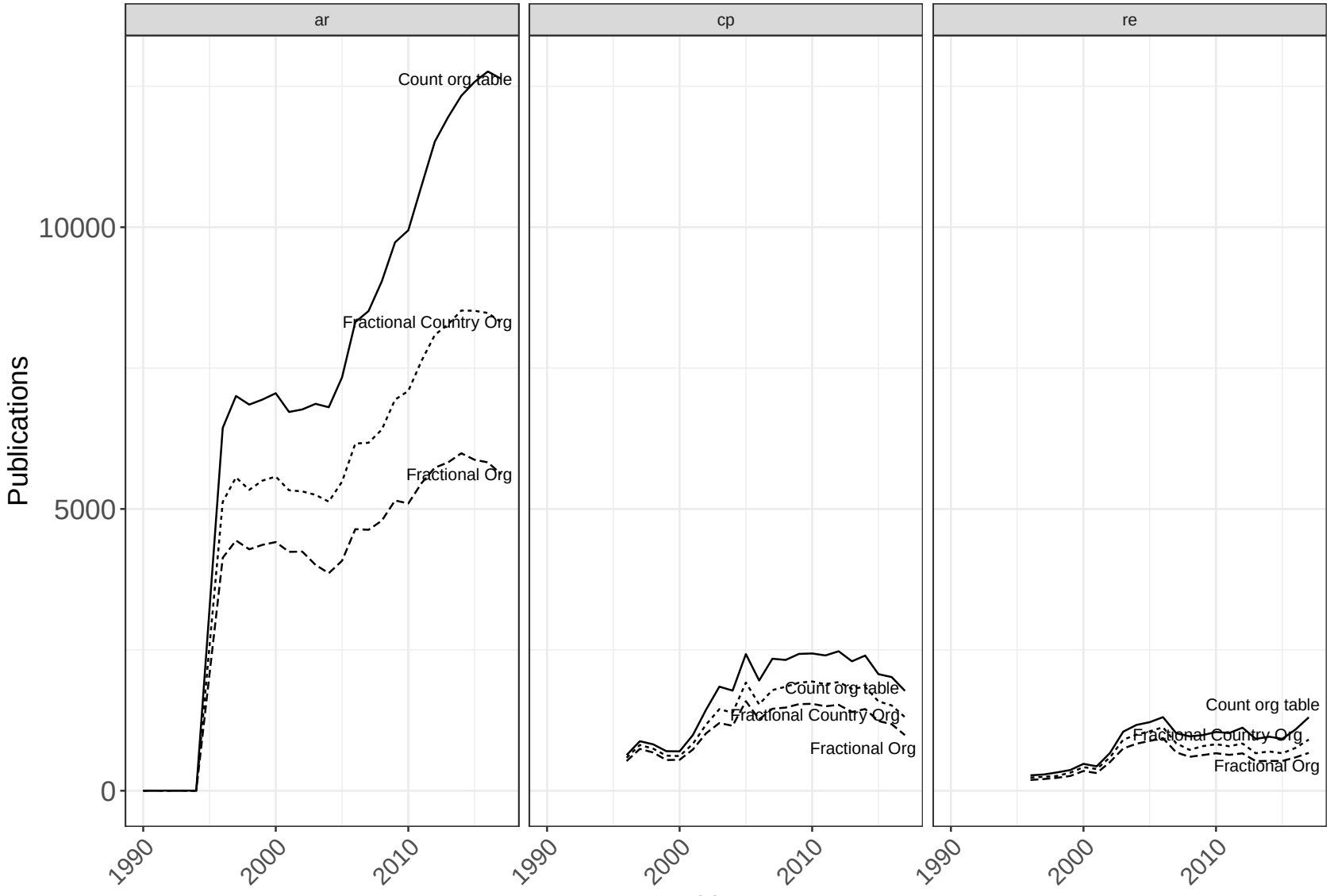


Figure 2: Articles, Reviews and Conference proceedings published by German institutes and universities in WOS and Scopus in 1990 - 2017

Berlin publications (Scopus)



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 geographical coordinates (4,723,171 items) were used

Unique organizations (problematic?!)

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

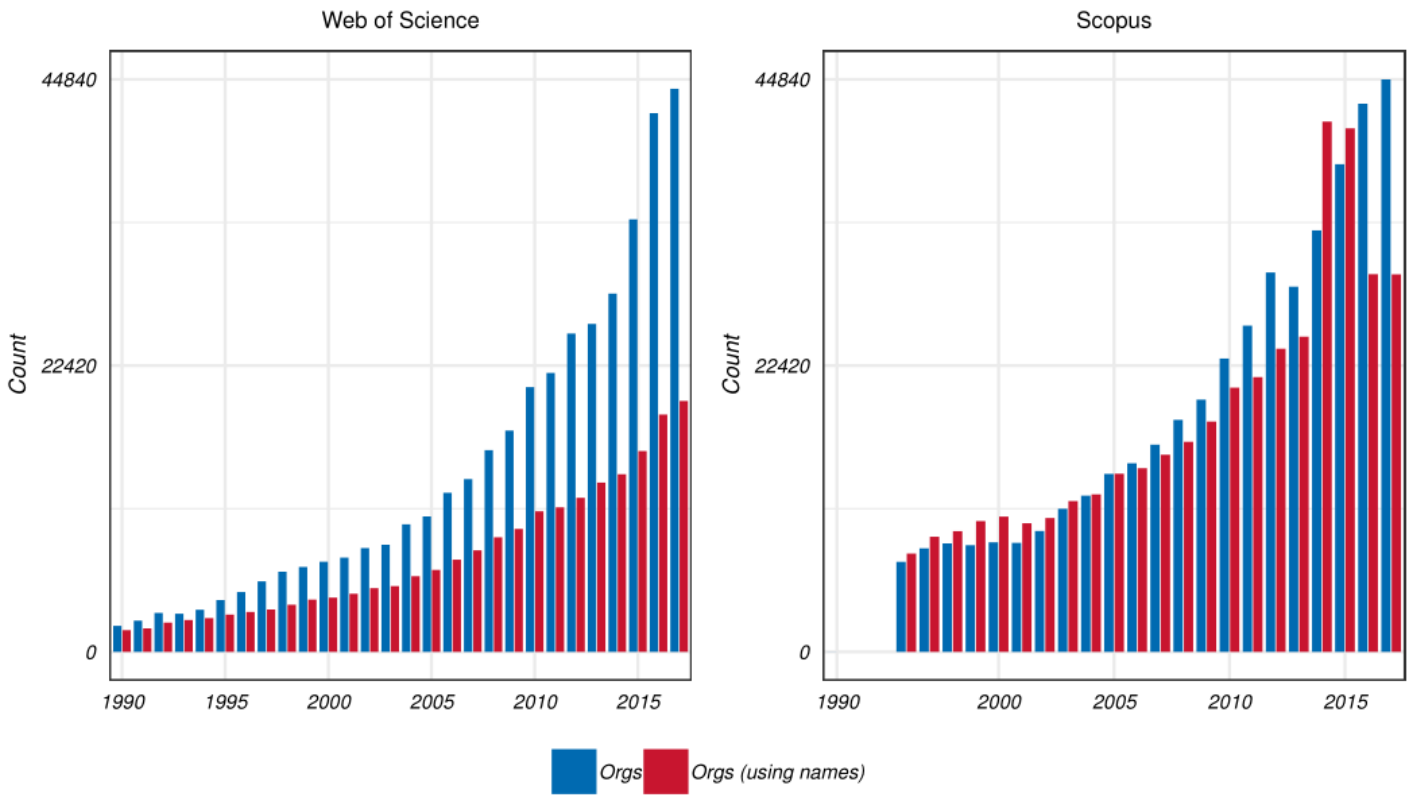


Figure 10: Unique organizations with which Berlin region institutes and universities have collaborated in Articles, Reviews and Conference proceedings in WOS and Scopus in 1990 - 2017

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
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

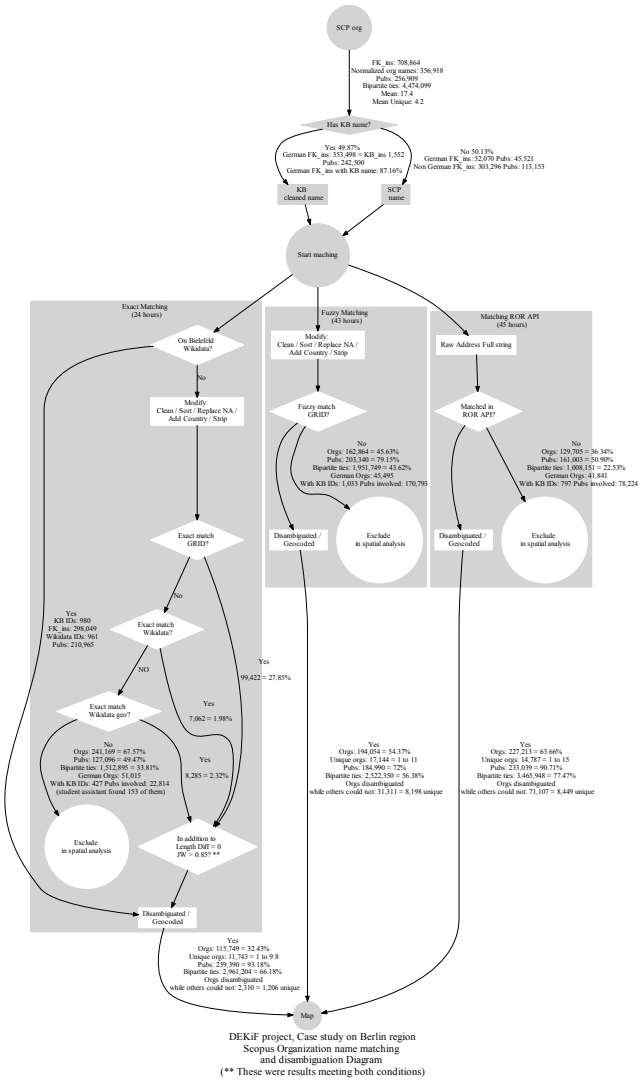
```
select PK_INSTITUTIONS, ORGANIZATION1, INSTITUTION_FULL from scopus_b_2018.institutions
where pk_institutions in ('10897353', '29324169', '14472460', '46335179', '31666249', '696390', '40120836','27118260', '28089912',
```

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 & geo-coding logic & results



Comparison of different disambiguation methods

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

Descriptive figures on map and sectors

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

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

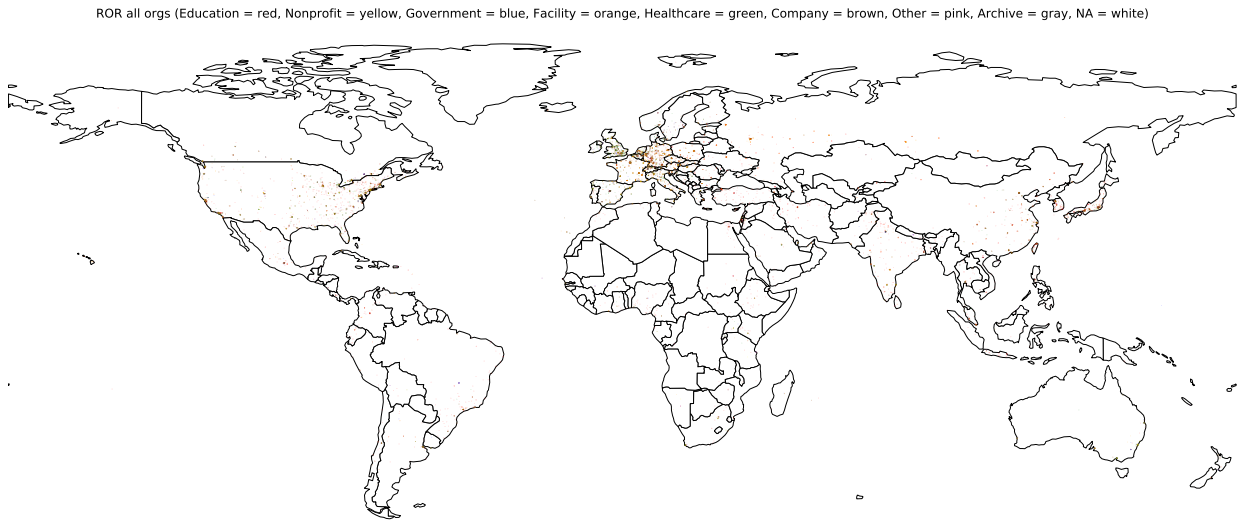
DZHW

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 of All disciplines



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

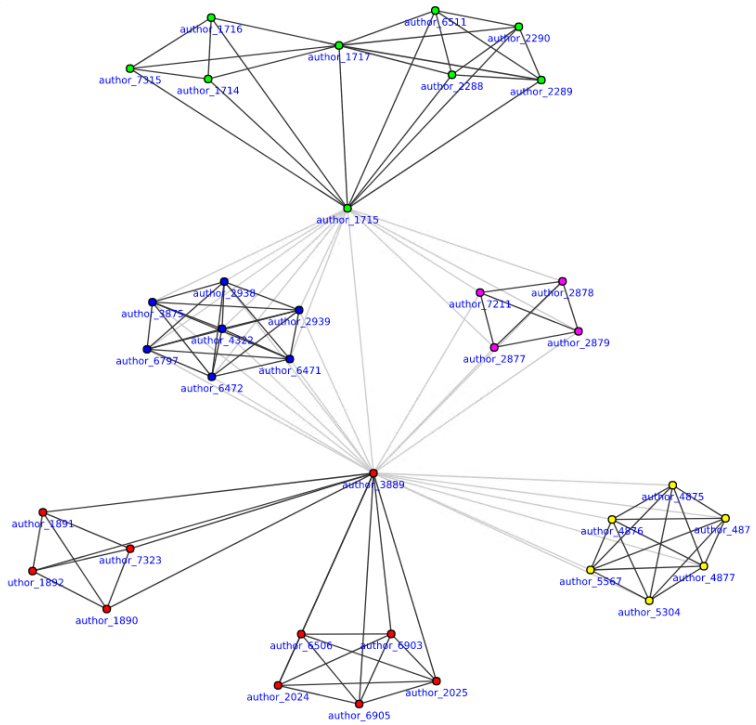
DZHW

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

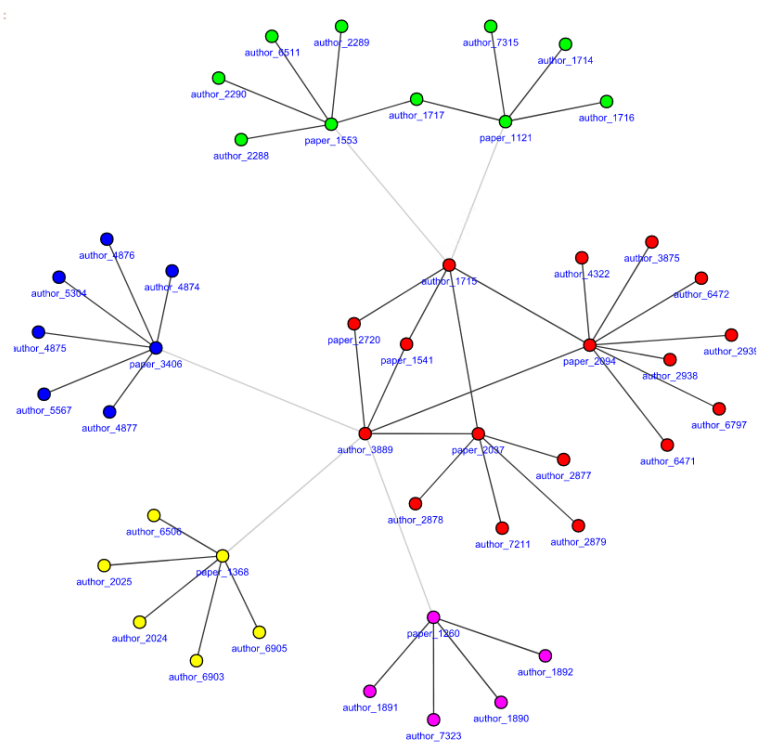
Bipartite community detection (1/2)

DZHW

[123]:



[26]:



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

Bipartite community detection (2/2)

DZHW

- Constant Pots Model in `Leidenalg` library
- Emphasizes the importance of links *within* communities vs. *between*
- Communities such that the link density between the communities (external density) is lower than γ and the link density within communities (internal density) is higher than γ
- $\gamma = 3 \times 10^{-4}$
- Yield *6,088* communities with a uniform distribution of org/pubs

cluster	type2	Count org/papers
0	org	476
0	paper	183469
1	org	1469
1	paper	16952
2	org	136
2	paper	4763
3	org	189
3	paper	1248
4	org	485
4	paper	310

cluster	COUNTRYCODE	Count org
0	DEU	193
1	DEU	27
2	DEU	18
3	DEU	19
4	DEU	11

Communities based on network structure (1/2)

DZHW

- One specific case, clusters 1 and 2 in Berlin G comp.
- Cluster 1: 16,952 papers / 1,469 orgs = 12 (avg)
- Cluster 2: 4,763 papers / 136 orgs = 35 (avg)

cluster	org_type	Count org	total_pubs	total_pubs_clu
1	'Archive'	22	1525	270085
1	'Company'	18	1167	270085
1	'Education'	734	206983	270085
1	'Facility'	327	38471	270085
1	'Government'	96	10065	270085
1	'Healthcare'	149	5027	270085
1	'Nonprofit'	87	4447	270085
1	'Other'	30	2323	270085
1	NA	6	77	270085
2	'Company'	14	251	9650
2	'Education'	50	2561	9650
2	'Facility'	55	6265	9650
2	'Government'	5	53	9650
2	'Healthcare'	3	96	9650
2	'Nonprofit'	5	280	9650
2	'Other'	4	144	9650

Countries in communities based on network structure (2/2)

cluster	COUNTRYCODE	Count org	cluster	COUNTRYCODE	Count org
1	USA	344	2	RUS	21
1	GBR	105	2	DEU	18
1	JPN	88	2	FRA	18
1	FRA	81	2	USA	11
1	AUS	70	2	CHN	10
1	ITA	67	2	ESP	5
1	ESP	62	2	KOR	5
1	CAN	45	2	POL	4
1	CHN	33	2	CAN	3
1	RUS	32	2	JPN	3
1	DEU	27	2	UKR	3
1	IND	24	2	UZB	3
1	KOR	20	2	BEL	2
1	SWE	20	2	BIH	2
1	TUR	20	2	BLR	2

Limitations

- Success stories only!
- No view on conflict in collaborations
- No view on unfinished works or unsuccessful collaboration
- No view on the motivations behind collaborations

Summary

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

Summary

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

Summary

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

Brief on Bibliometric case study on iDIV

Possible delineation scenarios for iDIV pubs

DZHW

- **Each answer a different research question**
- Take all publications of **11 main members** and limit to iDIV (*already done*)
- Take all publications of cities of those **11 main members** and limit to iDIV
- Take all **bio-diversity field** in world and limit to iDIV
  - Starting from known *journals* of the field
  - Starting from *themes and keyword* search
- Our selected approach to be most feasible:
  - *Starting from iDIV pub list online*, take all publications of members/scientists from Scopus (during their career) and then limit to biodiversity
  - *Starting from iDIV pub list online*, extract disciplines of them on Scopus, take all those disciplines and position iDIV in them

Discrepancies between data sources

DZHW

iDiv _ Publication List Analysis

Part III KB database

Search strategy (a)

- **Field:** DOIs extracted from iDiv's publication list
- **Data source:** WoS & Scopus _ B_2019

Data (a)

Table 7. Records by source

PY	iDiv_Publist		Web-Version		KB_WoS_2019		KB_SCP_2019	
	Counts	DOI	Wos	Scopus	Counts	DOI	Counts	DOI
2013	40	37	32	33	26	26	29	29
2014	179	171	154	164	139	139	155	155
2015	277	264	246	249	216	216	230	230
2016	359	352	335	341	321	321	327	327
2017	445	442	420	425	400	400	414	414
2018	386	378	350	353	338	338	345	345
2019	368	367	328	333	132	132	174	174
SUM	2054	2011	1865	1899	1572	1572	1674	1674

Explanation _Table 7

1. Data obtained by doi are less compared to others mentioned above, which might be easily understood because publications in 2019 are not fully covered by our KB database
2. In this case, SCP_b_2019 has more matched documents in comparison with WOS_b_2019
3. Data extracted from KB database herein are not limited to journal articles, conference proceedings and review

14

What happens next?

- Further analysis on co-authorship view to scientific collaboration (e.g., homophily effects, highest effects in probability of collaboration ties)
  - comparing community compositions in different organizational scenarios
- In order to not reinvent the wheel:
  - We are negotiating data sharing on R&R project collaborations (based on Forderkatalogue)
  - We are negotiating data sharing on patent collaborations (based on different sources)
- Multiplex view to these different forms of scientific collaboration

Thanks for your attention!