Comparative Analysis of Bibliometric weighting methods:

A Case Study of the University of Kassel and the Klinikum Kassel.

By: Joshua Adu

Outline

1. Objectives
2. Main Questions (Motivation)
3. Background
4. Data and Methodology
5. Data Processing
6. Results and Visualisations
7. Discussions

Objectives

Examine how different weighting methods affect the measured publication volume of the University of Kassel and Klinikum Kassel for 2023 publications.

University of Kassel: A full-fledged Public University in the German State of Hessen.

Klinikum Kassel: A Teaching Hospital in the German State of Hessen.

Main questions:

- Which method is preferred in a real bibliometric evaluation?
- Which factors (e.g., discipline, team size, authorship conventions) play a role in this decision?
- What differences arise between the methods?

Background

Bibliometrics – a statistical analysis of written publications, which offers essential tools for quantifying research output and impact (Pritchard, A.1969).

Importance: Foundational element in science policy, university and subject rankings, and allocation of funding.

Attribution problem: Determining a fair distribution of credit for a publication among its multiple co-authors and their affiliated institutions.

Full Counting (Traditional Method): Grants every participating institution a full (1) credit for a publication, irrespective of the number of co-authors.

Other Counting Methods: Other methods include, first author, harmonic counting, fractional counting and $1/\sqrt{N}$. These methods shall be tackled later

Data and Methodology

Source: OpenAlex database.

The raw data is structured into four tables:

- kassel2526_items: Which contains all Kassel publications (article-level)
- kassel2526_items_authors: Provides author-level information for Kassel publications
- kassel2526_authors_affiliations: Contains all authors with a Kassel affiliation
- kassel2526_items_affiliations: Contains all publications with a Kassel affiliation

Data Overview

kassel2526_items_authors:
item_id
author_seq_nr
family_name
orcid
openalex_author_id

kassel2526_items:
item_id
pubyear
item_title
source_title
source_type
item_type
author_count
keyword
class_name
first_page
last_page
cit_3_years
cit_all_years

kassel2526_authors_affiliations:
item_id
author_seq_nr
aff_seq_nr
organization
suborganization
vendor_org_id
author_id

kassel2526_items_affiliations:
item_id
aff_seq_nr
raw_affiliation_string
organization
suborganization
vendor_org_id
city
countrycode

RECORDS	130976	31307	65668	33586
COLUMNS	5	13	7	8

TOTAL	
1st Merge	130,956
2nd Merge	328,277
Duplicates	-20
	130956

UNIKASSEL VERSITÄT

Definition of Key Variables

Description of the relevant variables

- *item_id:* *unique publication identifier -> linked to other tables*
- *author_count:* *No. of contributing authors for an article/publication*
- *Author_count_new:* *No. of contributing authors from a specific institution for an article*
- *pubyear:* *year of publication*
- author_seq_nr: equals 1 if author is the first author for a publication/article
- *author_seq_nr:* authors position for each author
- organization: affiliation of author
- *item_title:* *title of the publication*
- *item_type:* *type of publication, e.g., article, review*
- *family_name:* *name of the author*
- *openalex_author_id:* *unique OpenAlex identifier each author*

Load and Merge Tables

Author-level information for Kassel publications

	item_id	author_seq_nr	family_name	openalex_author_id
2	W100059397	1	Martin Atzmueller	A5011835245
3	W100059397		Folke Mitzlaff	A5067244284
4	W1000991012		Maximilian Sch√§fer	A5014619402
5	W1000991012		Werner Thole	A5077488767
6	W1001237390	1	Frank Czerner	A5008700036
7	W100156544	1	H. Dieter Dahlhoff	A5078504152
8	W1002037175	1	Oskar Anderson	A5008442055

Authors with a Kassel affiliation

	item_id	author_seq_nr	aff_seq_nr	organization
2	W100059397	1		{"University of Kassel"}
3	W100059397			{"University of Kassel"}
4	W1000991012			{"University of Kassel"}
5	W1000991012			{"University of Kassel"}
6	W1001237390	1		{"University of Kassel"}
7	W100156544	1		{"University of Kassel"}
8	W1002037175	1		{"Institut f√ºr Angewandte Statistik"}

Kassel publications (article-level)

	item_id	pubyear	item_type	author_count
2	W100059397	2011	{article}	2
3	W1000991012	2014	{book-chapter}	3
4	W1001237390	2012	{book-chapter}	1
5	W100156544	2013	{book-chapter}	2
6	W1002037175	1976	{book-chapter}	5

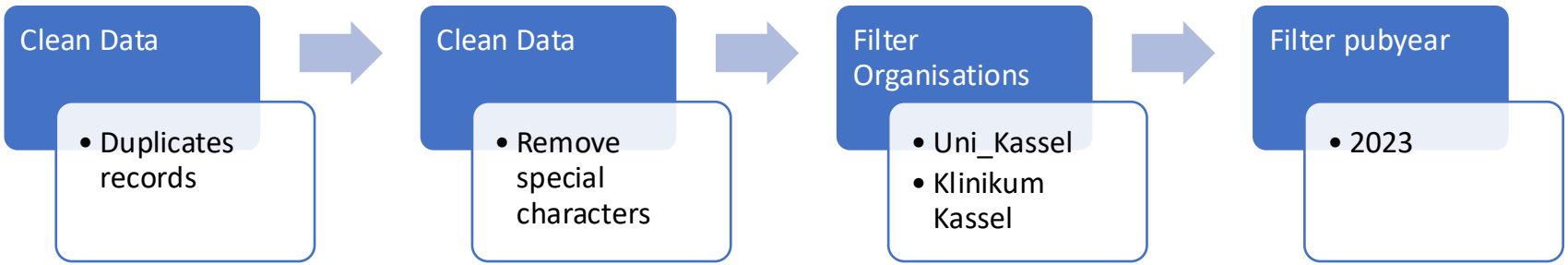
Remove duplicates

Merge using item_ID

UNIKASSEL VERSITÄT

Data Preparation

	item_id	author_seq_nr	pubyear	author_coun	aff_seq_nr	organization	openalex_author_id	author_id
2	W100059397	1	2011	2	NA	{"University of Kassel"}	A5011835245	A5011835245
3	W100059397	NA	2011	2	NA	{"University of Kassel"}	A5067244284	A5067244284
4	W1000991012	NA	2014	3	NA	{"University of Kassel"}	A5014619402	A5014619402
5	W1000991012	NA	2014	3	NA	{"University of Kassel"}	A5014619402	A5077488767
6	W1000991012	NA	2014	3	NA	{"University of Kassel"}	A5077488767	A5014619402
7	W1000991012	NA	2014	3	NA	{"University of Kassel"}	A5077488767	A5077488767
8	W1001237390	1	2012	1	NA	{"University of Kassel"}	A5008700036	A5008700036
9	W100156544	1	2013	2	NA	{"University of Kassel"}	A5078504152	A5078504152
10	W1002037175	1	1976	5	NA	{"Institut f√ºr Angewandte	A5008442055	A5008442055



Data Preparation cont….

	item_id	author_seq_nr	pubyear	author_count	organization
2	W2899097361	1	2023	2	uni_kassel
3	W2899097361	NA	2023	2	uni_kassel
4	W2969643344	1	2023	2	uni_kassel
5	W2969643344	NA	2023	2	uni_kassel
6	W2975543265	1	2023	1	uni_kassel
7	W3112822235	1	2023	2	uni_kassel
8	W3125862763	1	2023	3	uni_kassel
9	W3125862763	NA	2023	3	uni_kassel
10	W3125862763	NA	2023	3	uni_kassel

Create a Unique Column (item_id_org)
- Combine item_id and organization

Group by item_id_org
- Save count as author_count_new

Update author_seg_nr
- Uni_Kassel
- Klinikum Kassel

Create authorship type
- Single
- Full
- Co-author
- First Co-author

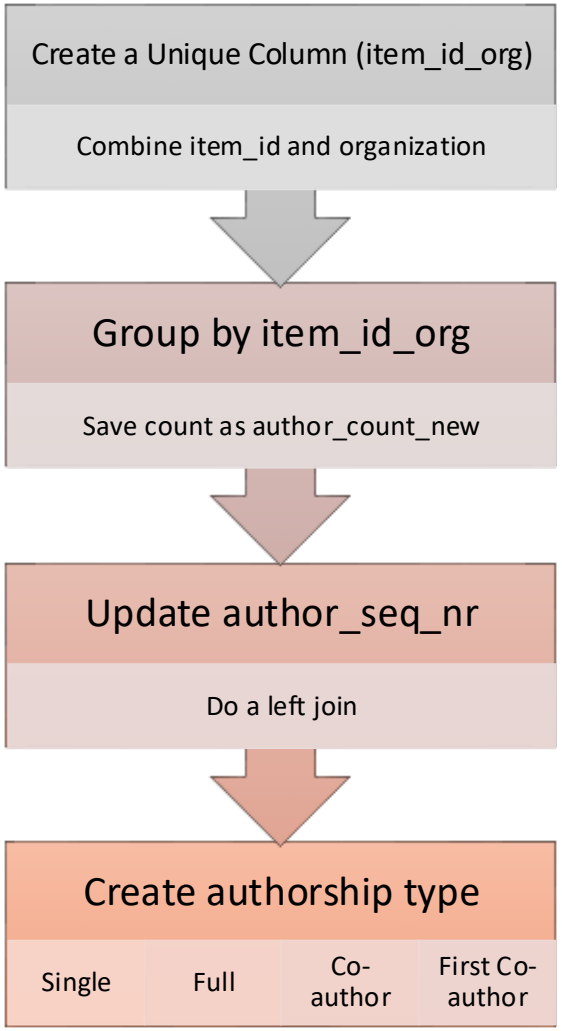
Data Preparation cont…..

Before

	item_id	author_seq_nr	pubyear	author_count	organization
2	W2899097361	1	2023	2	uni_kassel
3	W2899097361	NA	2023	2	uni_kassel
4	W2969643344	1	2023	2	uni_kassel
5	W2969643344	NA	2023	2	uni_kassel
6	W2975543265	1	2023	1	uni_kassel
7	W3112822235	1	2023	2	uni_kassel
8	W3125862763	1	2023	3	uni_kassel
9	W3125862763	NA	2023	3	uni_kassel
10	W3125862763	NA	2023	3	uni_kassel

After

	item_id	organization	author_count	author_count_new	author_seq_nr	authorship_type
2	W2899097361	uni_kassel	2	2	1	full
3	W2969643344	uni_kassel	2	2	1	full
4	W2975543265	uni_kassel	1	1	1	single
5	W3112822235	uni_kassel	2	1	1	first_co-author
6	W3125862763	uni_kassel	3	3	1	full



Full Counting Method

Each participating institution receives a score of 1, irrespective of the number of authors contributed by the institution

	item_id	organization	author_count	author_count_new	author_seq_nr	authorship_type	full_count
9	W3176204094	uni_kassel	1	1	1	single	1
10	W3182545621	uni_kassel	4	3	0	co-author	1
11	W3193739082	uni_kassel	5	5	1	full	1
12	W3202234318	uni_kassel	4	4	1	full	1
13	W3210286092	uni_kassel	4	1	1	first_co-author	1
14	W3211027958	uni_kassel	2	2	1	full	1
15	W4221074316	uni_kassel	2	2	1	full	1
16	W4223540547	kli_kassel	3	3	1	full	1
17	W4229011028	uni_kassel	9	8	0	co-author	1
18	W4285162186	uni_kassel	90	89	0	co-author	1

Full_count

- Assign 1 to each participating institution

UNIKASSEL VERSITÄT

Fractional Counting Method

Each publication is distributed equally across all contributing authors: Wi = i/N. Fractional count = W_i* number of authors from institution.

Example: A publication with five authors contributes 0.2 to the publication volume of each. 1 publication with 4 authors → each receives 0.25.

	item_id	organization	author_count	author_count_new	author_seq_nr	authorship_type	full_count	frac_count
9	W3176204094	uni_kassel	1	1	1	single	1	1.00
10	W3182545621	uni_kassel	4	3	0	co-author	1	0.75
11	W3193739082	uni_kassel	5	5	1	full	1	1.00
12	W3202234318	uni_kassel	4	4	1	full	1	1.00
13	W3210286092	uni_kassel	4	1	1	first_co-author	1	0.25
14	W3211027958	uni_kassel	2	2	1	full	1	1.00
15	W4221074316	uni_kassel	2	2	1	full	1	1.00
16	W4223540547	kli_kassel	3	3	1	full	1	1.00
17	W4229011028	uni_kassel	9	8	0	co-author	1	0.89
18	W4285162186	uni_kassel	90	89	0	co-author	1	0.99

Frac_count

- W_i*number of authors from institution

First Author Weighting

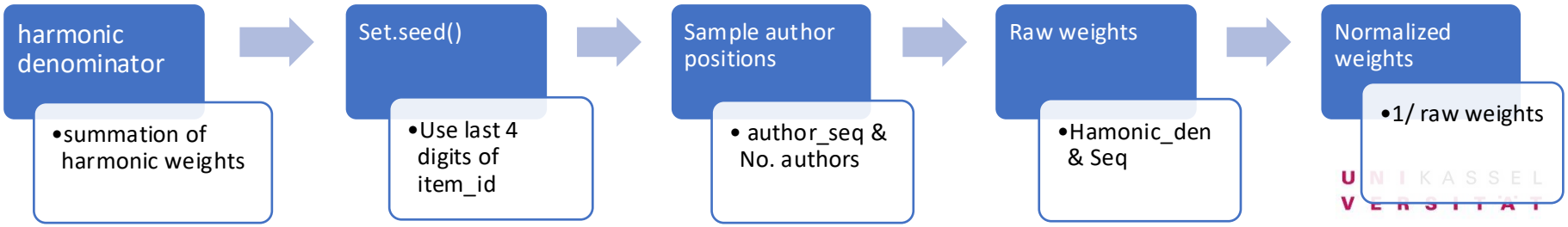
The first author receives 0.5, while the remaining 0.5 is distributed equally among the other authors ($W_i = 0.5/(n-1)$ for I = 2,….,n.

	item_id	organization	author_count	author_count_new	author_seq_nr	authorship_type	full_count	frac_count	first_author_count
9	W3176204094	uni_kassel	1	1	1	single	1	1.00	1.0
10	W3182545621	uni_kassel	4	3	0	co-author	1	0.75	0.5
11	W3193739082	uni_kassel	5	5	1	full	1	1.00	1.0
12	W3202234318	uni_kassel	4	4	1	full	1	1.00	1.0
13	W3210286092	uni_kassel	4	1	1	first_co-author	1	0.25	0.5
14	W3211027958	uni_kassel	2	2	1	full	1	1.00	1.0
15	W4221074316	uni_kassel	2	2	1	full	1	1.00	1.0
16	W4223540547	kli_kassel	3	3	1	full	1	1.00	1.0
17	W4229011028	uni_kassel	9	8	0	co-author	1	0.89	0.5
18	W4285162186	uni_kassel	90	89	0	co-author	1	0.99	0.5

Harmonic Counting

- Shares decrease harmonically with the author position
- Weights are calculated according to the author's position and the total number of authors.
- Example: publication with 3 authors.
- Raw weights: 1/1 (author 1), 1/2 (author 2), 1/3 (author 3). Sum = 1 + 0.5 + 0.333… = 1.833.
- Normalized weights: author 1 = 1 / 1.833 ≈ 0.545, author 2 = 0.5 / 1.833 ≈ 0.273, author 3 = 0.333 / 1.833 ≈ 0.182.

NB: author_seq -> a randomly generated authorship position for each author



Harmonic Counting

	item_id	organization	author_count	author_count_new	author_seq_nr	authorship_type	full_count	frac_count	first_author_count	harmonic_count
8	W3171354287	uni_kassel	4	4	1	full	1	1.00	1.0	1.00
9	W3176204094	uni_kassel	1	1	1	single	1	1.00	1.0	1.00
10	W3182545621	uni_kassel	4	3	0	co-author	1	0.75	0.5	0.52
11	W3193739082	uni_kassel	5	5	1	full	1	1.00	1.0	1.00
12	W3202234318	uni_kassel	4	4	1	full	1	1.00	1.0	1.00
13	W3210286092	uni_kassel	4	1	1	first_co-author	1	0.25	0.5	0.48
14	W3211027958	uni_kassel	2	2	1	full	1	1.00	1.0	1.00
15	W4221074316	uni_kassel	2	2	1	full	1	1.00	1.0	1.00
16	W4223540547	kli_kassel	3	3	1	full	1	1.00	1.0	1.00
17	W4229011028	uni_kassel	9	8	0	co-author	1	0.89	0.5	0.65
18	W4285162186	uni_kassel	90	89	0	co-author	1	0.99	0.5	0.70
19	W4285245056	uni_kassel	4	3	0	co-author	1	0.75	0.5	0.52

1/√N Counting

- Each contributing author receives **1**/√N.
- Each institution receives ((**1**/√N)*n). Where small n is No. of Authors from Institution.

Example: 1 publication with 4 authors → each receives 0.5 (**1**/√4)instead of 0.25.

	item_id	organization	author_count	author_count_new	author_seq_nr	authorship_type	full_count	frac_count	first_author_count	harmonic_count	inv_sqrt_count
8	W3171354287	uni_kassel	4	4	1	full	1	1.00	1.0	1.00	2.00
9	W3176204094	uni_kassel	1	1	1	single	1	1.00	1.0	1.00	1.00
10	W3182545621	uni_kassel	4	3	0	co-author	1	0.75	0.5	0.52	1.50
11	W3193739082	uni_kassel	5	5	1	full	1	1.00	1.0	1.00	2.24
12	W3202234318	uni_kassel	4	4	1	full	1	1.00	1.0	1.00	2.00
13	W3210286092	uni_kassel	4	1	1	first_co-author	1	0.25	0.5	0.48	0.50
14	W3211027958	uni_kassel	2	2	1	full	1	1.00	1.0	1.00	1.41
15	W4221074316	uni_kassel	2	2	1	full	1	1.00	1.0	1.00	1.41
16	W4223540547	kli_kassel	3	3	1	full	1	1.00	1.0	1.00	1.73
17	W4229011028	uni_kassel	9	8	0	co-author	1	0.89	0.5	0.65	2.67
18	W4285162186	uni_kassel	90	89	0	co-author	1	0.99	0.5	0.70	9.38
19	W4285245056	uni_kassel	4	3	0	co-author	1	0.75	0.5	0.52	1.50

Results and Visualisation

No. of Authors from Kassel Institution

author_count_new

7,237

Key Insights: Klinikum Kassel published significantly less than the University of Kassel (about 39 times more).
This substantial difference is likely due to variances in institutional focus, faculty size, and funding sources.

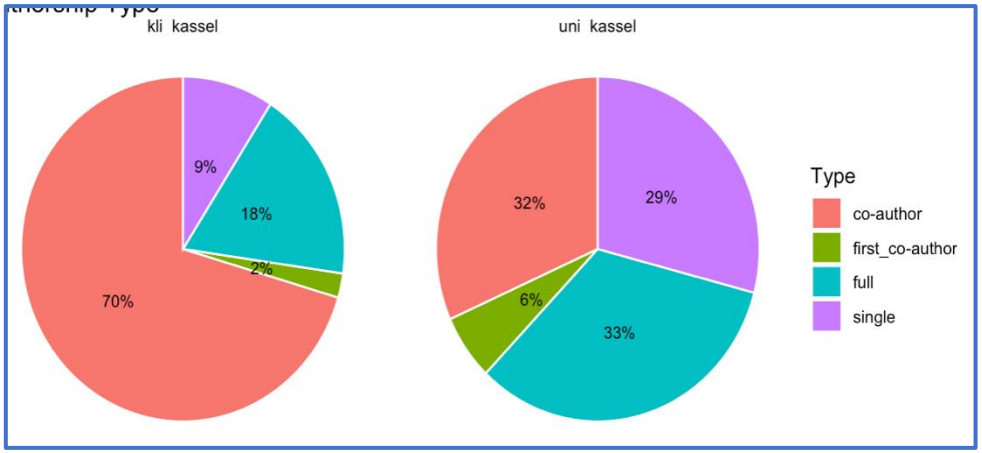
Publications Per Organisation

	organization	item_id
1.	uni_kassel	1,692
2.	kli_kassel	44

Key Insights: Klinikum contributed more authors 720 on 44 articles relative to the 6,517 to 1,692 of the University Kassel.

Reflecting the relatively high level of collaborations at the Klinikum Kassel.

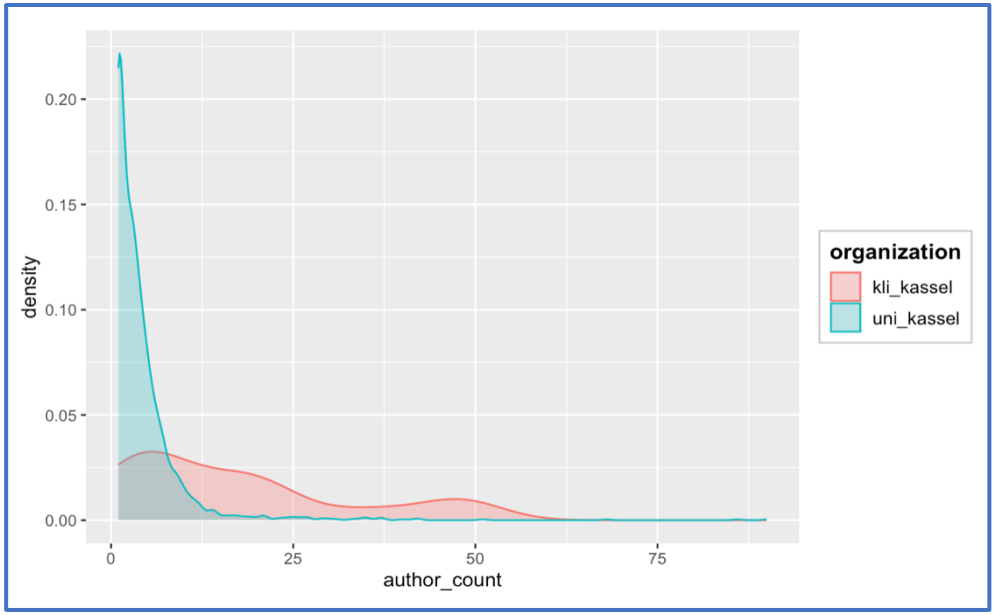
Authorship Type



Key Insights: Klinikum Kassel demonstrated a significantly higher rate of co-authorship 94% Vs 70% of Uni-Kassel.

Conversely, Uni-Kassel had a higher percentage of publications co-authored with internal members (33%) Vs Klinikum Kassel (18%).

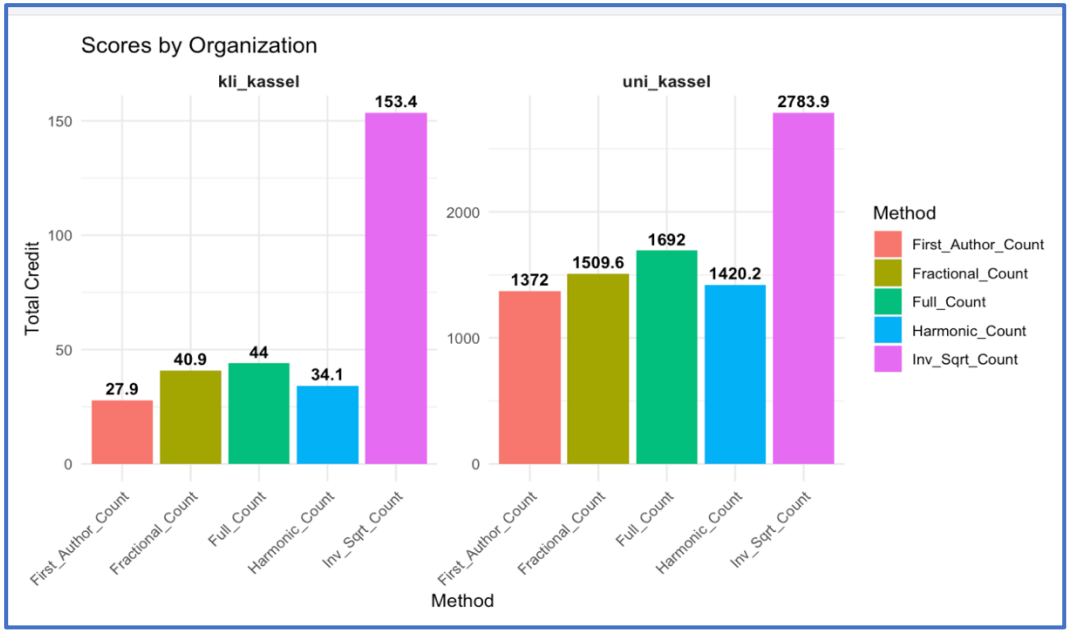
Density Plot of Author Count



Key Insights: The density plot highlights majority of publications feature an author count centered around 3 to 4 co-authors for Uni-Kassel Vs more dispersed distribution seen at Klinikum Kassel.

UNIKASSEL VERSITÄT

Absolute Scores



Key Insights: Based on the raw scores, Uni-Kassel out -performed on all metrics.
But does not reflect the reality due to faculty sizes.

Score per Counting Method

	organization ▲	first_author_count	full_count	frac_count	harmonic_count	inv_sqrt_count
1..	kli_kassel	27.92	44	40.9	34.12	153.4
2..	uni_kassel	1,372	1,692	1,509.6	1,420.17	2,783.87

Modelling with Hypothetical data

Key Insights:

- This Section uses hypothetical data to Create models under 3 assumptions.
- Models are used to explain behaviour of each counting method.
- Data was generated using doubling geometric series (2^(0:(n-1))). Where N = 10.

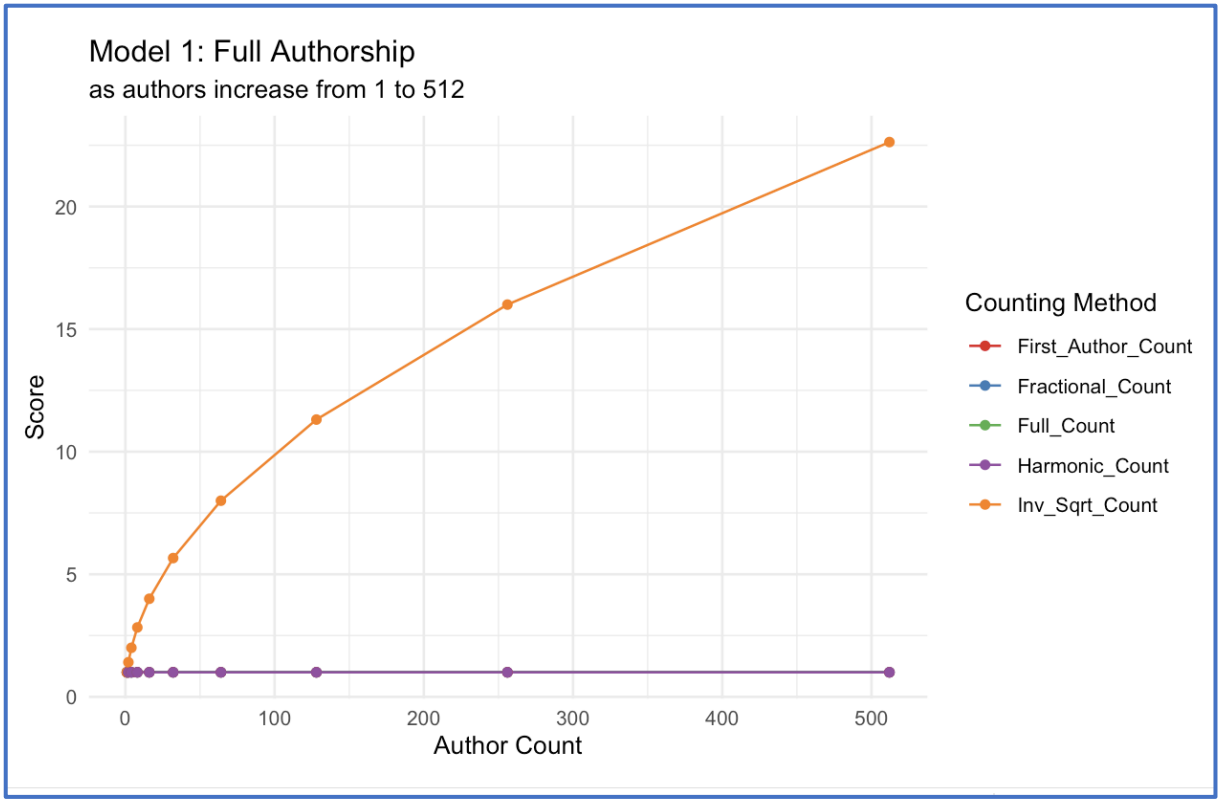
Models and Assumptions:

a) Model 1: full authorship – All authors are from the same institution

b) Model 2: Co-Authorship with Only the 1st Author as External

c) Model 3: Co-Authorship with only the 1st Author from the institution

Full Authorship

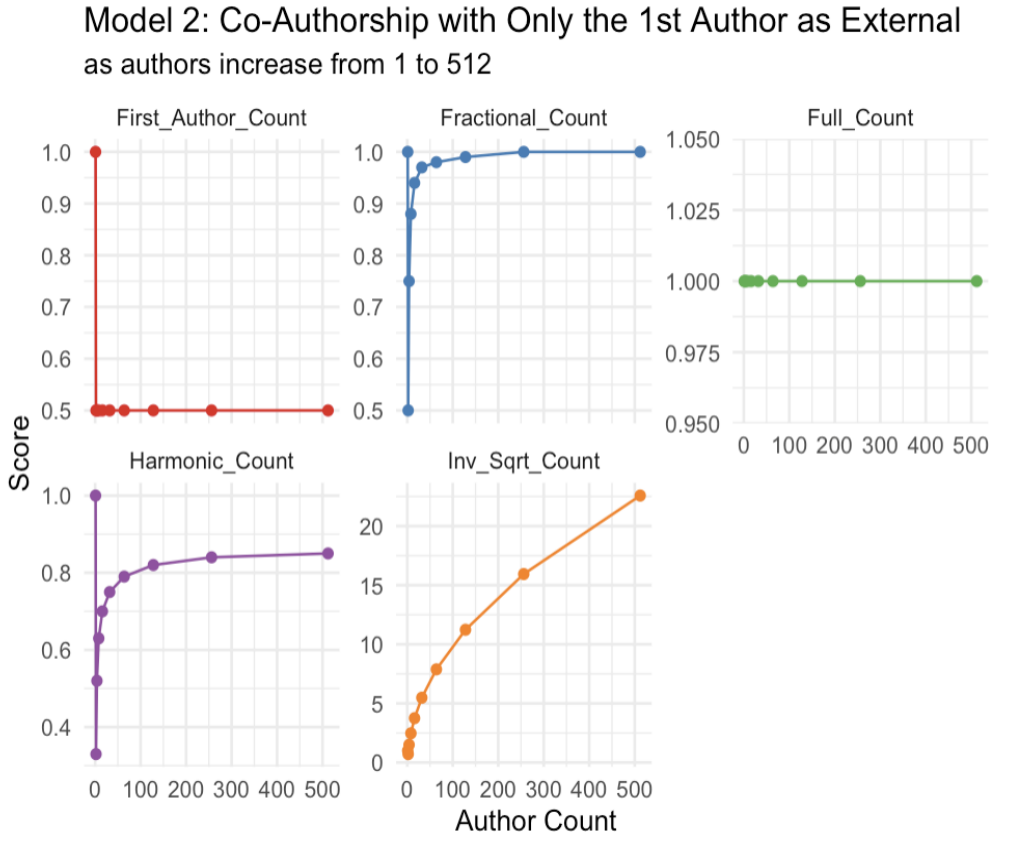


Key Insights:

- **Team size does not matter**
- **Position does not matter**

Exception: $1/\sqrt{N}$ method, high N implies high score

Co-Authorship with only the 1st Author as External



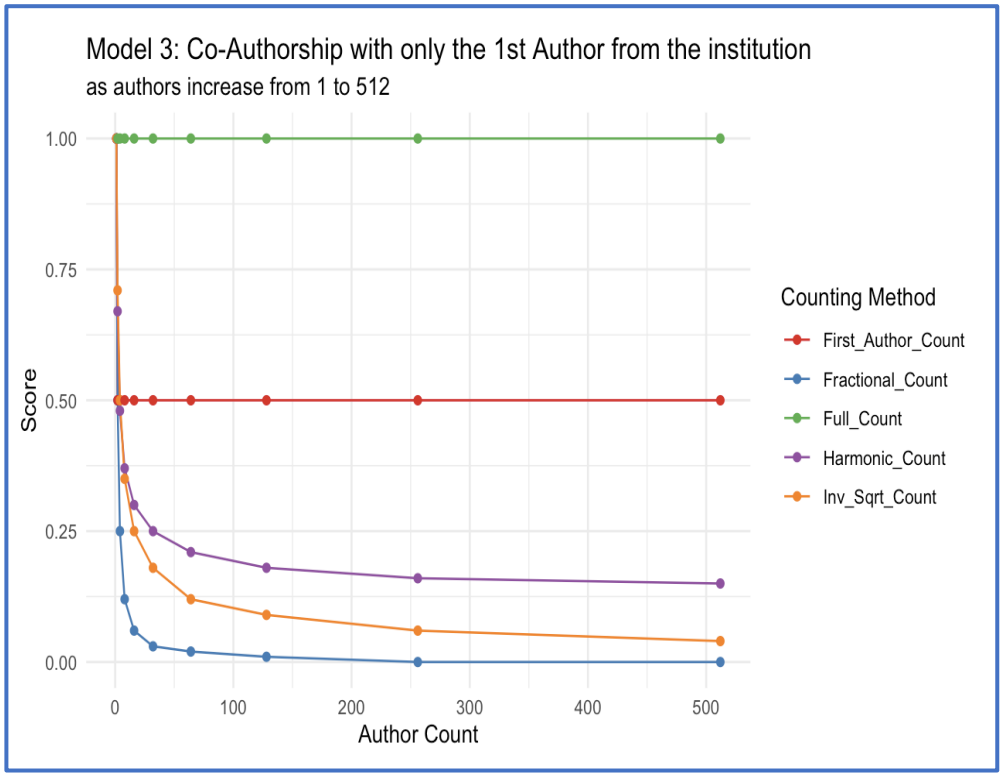
Key Insights:

- Fractional Counting: rises sharp but plateaus at 1 as $N \to \infty$.
- Harmonic Counting: rises but then increases at a decreasing rate as $N \to \infty$.
- $1/\sqrt{N}$: Score increases as n increases

Key Insights:

- Team size matters for Fractional and Harmonic Counting.
- Harmonic manages team size and posision dynamics better

Co-Authorship with only the 1st Author from the institution



Key Insights:

- Fractional Counting: sharp fall but plateaus at 0 as $N \rightarrow \infty$.
- Harmonic Counting: Decline but then decreases at a decreasing rate as n grows lager increases.
- Inverse Square Root: decreases as n increases

Key Insights:

- Position matters for Fractional and Harmonic Counting.
- Harmonic manages team size and posision dynamics better

UNIKASSEL VERSITÄT

Discussions

Which method is preferred in a real bibliometric evaluation?

Full Counting

- Promotes collaboration amongst institutions
- Does not factor team size and author sequence
- Easy to use and justify

Fractional Counting

- Contributions acknowledge equally
- Does factor team size
- Easy to use and justify than harmonic counting

First Author

- Difficult to justify first author share
- Does factor team size and author sequence
- Easy to use

Harmonic Count

- First author share varies with size of team
- Values lies between full count and first author
- Promotes collaboration amongst institutions and large teams
- Difficult to compute and justify

Discussions

Which method is preferred in a real bibliometric evaluation?

The Harmonic Counting method offers a compelling middle ground. Unlike Fractional counting, which treats a first author and a middle muthor as equal contributors (both getting 1/N).

Harmonic counting rewards the intellectual leadership associated with the ranks whiles considering the size of the team. Unlike First Author counting which which assigns 0.5 to the first author regardless of the team size.

Discussions

Which factors (e.g., discipline, team size, authorship conventions) play a role in this decision?

- Disciplinary differences: Some disciplines naturally requires collaborations, large group sizes, clinical trials, regulatory approvals.
- Team size: affects first author, fractional, harmonic methods.
- Authorship conventions: Where author sequence follows an alphabetical convention instead of reflecting contribution level.

What differences arise between the methods?

- Team size
- Authorship sequence
- Institutional collaborations

Thank you

Joshua Adu