International bibliometric benchmark
Institute for Logic, Language and Computation, University of Amsterdam
New mission-tailored approach in support of SEP evaluation 2024

Kasper Abcouwer | Bibliometric Support Office | University Library | University of Amsterdam
Laura Compier | Executive Office, Faculty of Science| University of Amsterdam
Manuel Llano Martínez| Bibliometric Support Office | University Library | University of Amsterdam

Executive summary

We here present the results obtained with a novel approach based on OpenAlex's Topics[1] classification. By pooling the Institute for Logic, Language and Computation (ILLC)'s major Topics into a custom Research Area, a fingerprint of the institute's academic 'niche' has been obtained, further referred to as the **ILLC Core Area**. A range of bibliometric indicators for ILLC and UvA have subsequently been compared with those of the world's most active institutions in the ILLC Core Area. This has resulted in a benchmark, tailored to the disciplinary main orientation of ILLC, as required by the Dutch Strategy Evaluation Protocol 2021-2027 (SEP).

This document comprises the following sections:

1. Key findings
2. Methodology and data collection
3. Analysis: ILLC output 2018-2023 and its Core Area
4. Analysis: Global trends for the ILLC Core Area
5. Conclusions

We show that ILLC's 15 largest Topics provide a representative selection of 56% of its OpenAlex publications, reflecting its main research foci and covering all six research units. OpenAlex coverage of ILLC's (peer refereed) articles and reviews is as high as 100%.

Among institutions in the ILLC Core Area, UvA is ranked 8th globally in terms of paper count. In the ILLC Core Area top 12, UvA ranks eleventh for total citation count and UvA ranks eighth on both the 'crown indicators' Sub-Field Normalized Citation Impact (SFNCI) and Sub-Field Normalized Impact per author (SFNCI/auth) respectively.[2] Please refer to Table 5 and 6 for an overview of all metrics and benchmark institutions.

The ILLC institute demonstrates significant impact with an SFNCI score of 4.33, more than quadrupling the global average in its research area. Notably, each of the institute's six units exceeds the global average impact. However, there are variations in impact among the units. For instance, the SFNCI score for Mathematical & Computational Logic (MCL) is one and a half times the world average, whereas the score for Natural Language Processing & Digital Humanities (NLP&DH) is six times the world average. Refer to Table 3 for detailed comparisons.

Within the ILLC Core Area, there are roughly two main groups: Chinese institutes and Anglo-American institutes, with the University of Amsterdam (UvA) being strongly connected to the latter. This distinction is evident from the patterns of incoming and outgoing citations (Tables 1 and 2) and the most productive institutes in the Core Area (Tables 5 and 6). These two groups show limited connectivity through co-publications, as highlighted in Table 7. Generally, Anglo-American institutions exhibit a higher impact, as indicated by their SFNCI values in Table 6, compared to Chinese institutions. Collaborative publications tend to enhance impact, including those involving Chinese and Anglo-American partnerships. Within the ILLC Core Area, UvA predominantly collaborates with the Universities of Edinburgh and Oxford.

[1] In OpenAlex each publication with sufficient citations or references is linked to one of appr. 4500 *Topics*, representing collections of documents with a "common focused intellectual interest" (see: https://github.com/ourresearch/openalex-topic-classification). These are determined with a clustering algorithm based on extended direct citations, developed by the Leiden CWTS (Waltman & van Eck 2012 and Klavans & Boyack 2017). Unlike traditional journal classifications such as Elsevier's ASJC (with ±330 subcategories) or Clarivate's (±254) WoS research areas, *Topics* are completely independent of journal classification. An important advantage of this new classification is the lack of overlap: a paper is assigned to no more than 1 *Topic*, as opposed to the up to 13 ASJC labels that journals (and all their articles) may receive. In addition, *Topics* represent real disciplinary micro niches that tend to be well recognized by active authors as (one of their) own areas of expertise.

[2] SFNCI is a metric that normalize citation impacts by field. It provides an indication of how a researcher's publication citations compare to the average or expected number of citations received by similar publications. The similarity of publications is determined based on factors such as year, type, and discipline.

Key findings

- All research units in the ILLC perform over average in their respective fields, with the ILLC as a whole quadrupling the global impact average in its Core Area.
- Within its Core Area, the ILLC ranks 8th worldwide both in output and in field-normalized impact of their publications.
- The most frequent external collaborating institutions are the Universities of Edinburgh and Oxford.
- Internal collaboration among ILLC research units is common across the board, with FSPL being the most frequent collaborator.
- During the assessment period there was a steady increase in output in Natural Language Processing and Language Models, which reflects a worldwide trend. The NLP&DH unit is the most impactful (six times above average) and has the largest output.
- Leveraging OpenAlex and Scopus data in addition Pure, we were able to identify 567 additional research outputs not registered in Pure with a DOI.

Methodology and data collection

The objective of this document is to comparatively assess the impact of ILLC's research in the global stage. In order to do this, since the research foci of the ILLC are often spread out across different departments in other institutions, we will focus on a Core Area of topics that are consistently categorised on the basis of citation networks and natural language processing for all universities. This allows for a comparison to be possible and meaningful between the ILLC and outputs in universities around the world, but also between different research units within this multidisciplinary institute, which have different publication volume and research culture.

International Benchmark methodology

The following steps outline the methodology used to define the ILLC Core Area and benchmark its research performance against global institutions:

Extracting ILLC DOIs from PURE/CRIS and Supplementing with OpenAlex
The initial dataset consists of Digital Object Identifiers (DOIs) representing the publications from the Institute for Logic, Language, and Computation (ILLC) retrieved from the PURE/CRIS systems. Where necessary, this set is supplemented with additional DOIs from OpenAlex and Scopus to ensure a comprehensive collection of ILLC's research outputs.
Harvesting Metadata from OpenAlex
Once the DOIs have been collected, OpenAlex is used to harvest the associated metadata for each publication. This metadata includes the assignment of each publication to a primary topic, using OpenAlex's Topics classification system. The primary topic reflects the core intellectual area of focus for each publication.
Ranking Primary Topics and Defining the ILLC Core Area
The primary topics of the ILLC publications are ranked based on the number of publications associated with each topic. The largest topics—those comprising around 50% of the total number of publications—are selected to define the ILLC Core Area. This method ensures that the ILLC Core Area represents the main disciplinary orientations of the institute.
Collecting Publications for the ILLC Core Area (2018-2023)
For the defined ILLC Core Area, all relevant publications from the 2018-2023 period are collected. This

includes both the publications produced by ILLC and those from other institutions that are active in the same research topics, providing a comprehensive dataset for benchmarking.

Calculating the Field Normalized Citation Impact for Relevant Subfields

Each primary topic is part of a broader subfield. To normalize the citation impact across these subfields, all publications from OpenAlex for the relevant subfields during the 2018-2023 period are harvested. For each year and publication type within these subfields, a reference citation value is calculated. This value represents the Subfield Normalized Citation Impact (SFNCI) and serves as a baseline for comparing citation performance across different research areas. The SFNCI allows for fair comparison by adjusting for variations in citation practices and impact levels between subfields.

Comparing, Analyzing, and Creating the Core Area International Benchmark

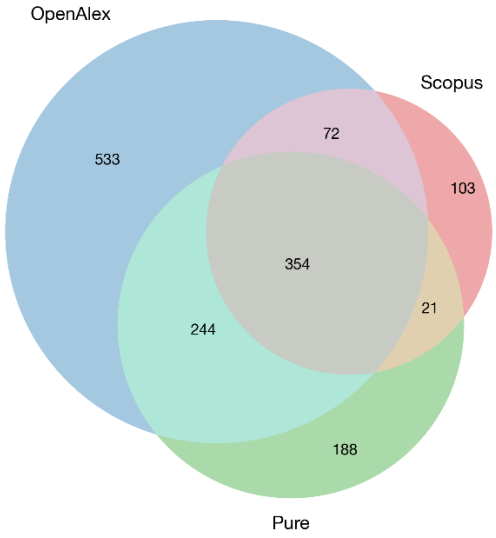
The final step involves comparing the bibliometric performance of ILLC and the University of Amsterdam (UvA) with other leading institutions within the ILLC Core Area. A range of bibliometric indicators—such as citation impact, publication volume, and the SFNCI—are applied to create an international benchmark. This benchmark is specifically tailored to reflect the ILLC's disciplinary orientation and meets the standards set by the Dutch Strategy Evaluation Protocol 2021-2027 (SEP).

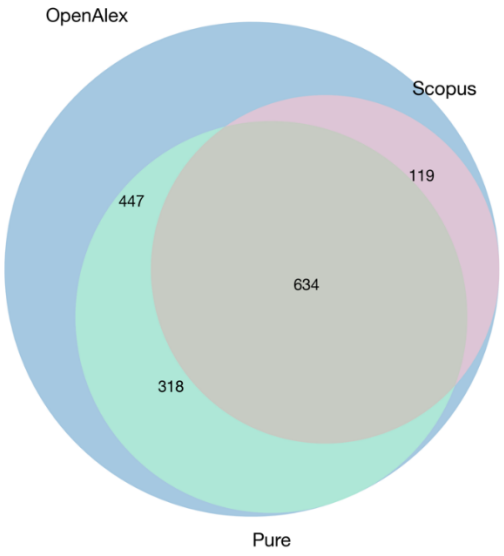
Data retrieval and processing

We created a dataset of 1,519 reportable publications by ILLC between 2018 and 2023, each with a unique Digital Object Identifier (DOI). The reportable research outputs correspond to the following SEP categories: refereed article, non-refereed article, conference paper, preprint, book, and book chapter. This dataset was compiled from three different sources: OurResearch's OpenAlex (an open bibliometric database), Elsevier's Scopus (a closed proprietary bibliometric database), and Elsevier's Pure (the University of Amsterdam's current research information system). By aggregating these sources and conducting manual checks, we aimed for a high level of completeness and accuracy, surpassing what is achieved by relying solely on publications registered in Pure.

Scopus and OpenAlex provide affiliation data only up to the university level, missing data on faculty or research institute affiliation. Thus, the best way to query relevant data is via author identifiers. The institute manager kindly provided a list of academic staff, their affiliations, and their appointment periods. We manually added Scopus and OpenAlex author identifiers to most of the staff (postdoctoral researchers, assistant professors, associate professors, full professors, and emeriti). All staff in these categories, except for three postdoctoral researchers, were identified in both databases, totalling 164 authors. We excluded PhD candidates and guest researchers due to time constraints, their often transient tenure, differences in output, and difficulties in finding reliable identifiers for early-career researchers. For these groups, we relied solely on publications registered in Pure.

Figure 1) left venn-diagram: overlap and provenance of the aggregated data sources ILLC publications 2018-2023; right venn-diagram: presence ILLC publications in Pure, Scopus and OpenAlex.





There are a total of 1,315 items attributed to the ILLC in Pure for the period 2018-2023, out of which 984 have a DOI, and 807 are reportable for the purpose of this analysis. These were aggregated with the data queried from OpenAlex and Scopus, which was manually checked in collaboration with the faculty executive officer to weed out errors and select only reportable outputs. This resulted in 708 additional items with a DOI that were found in Scopus and OpenAlex, making for a total of 1,519 unique reportable publications. The left side diagram in Figure 1 shows the provenance and overlap of the aggregated data sources.

Once the data was aggregated, all resulting DOIs were queried again in OpenAlex, Scopus, and the whole of Pure to determine how many of these outputs are present in each database, regardless of whether they were initially discovered following the method outlined above. The result is as follows: all 1,519 (100%) research outputs are present in OpenAlex, 952 (62%) are present in Pure (in other words, 145 additional publications were registered in Pure but not correctly attributed to the ILLC, and thus absent from the initial report), and 753 (49%) are present in Scopus (see Figure 1, right).

Overall, with this approach, we were able to find 567 additional research outputs not registered in Pure with a DOI (334 preprints, 101 refereed articles, 92 conference papers, 34 book chapters, 4 books, and 3 non-refereed articles).

Analysis

We analyzed the ILLC output for the period 2018-2023 at two levels: the complete set of publications and the Core Area set. The entire ILLC output was further analyzed at the unit level. The ILLC Core Area was examined as an independent set and then benchmarked globally against the most productive institutions in the world within this research area.

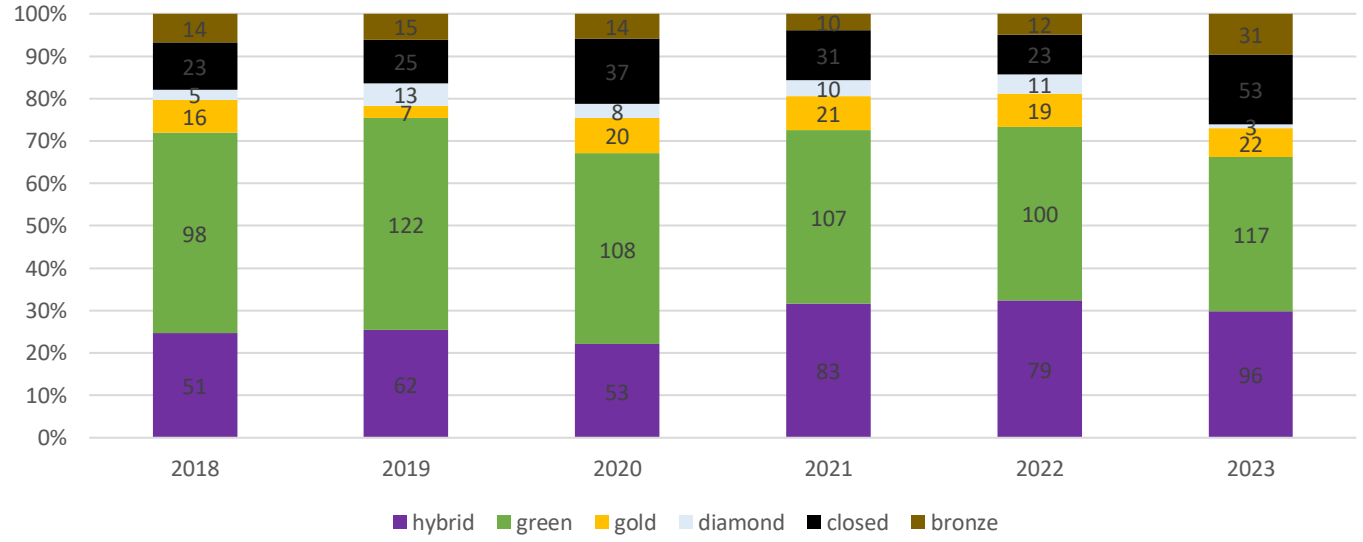
ILLC output 2018-2023 and its Core Area

Open access trends

Open Access is commonly classified according to a colour code. Green open access works are available in open repositories or archives and can sometimes also be published in toll-access journals. Hybrid works are made available free of charge to readers at the time of publication after the author typically pays the publisher an article processing charge (APC). These hybrid works are published in journals that include both open and closed articles. Gold open access works share the same characteristics as hybrid works but are published in journals that exclusively publish open access content. Diamond open access works are a type of gold open access where the publisher does not charge any article processing charges. Bronze open works are accessible to readers free of charge without an open license, though they are often not immediately open after publication and the publisher retains the right to remove them unilaterally and are thus not considered open access proper. Lastly, closed access works require a subscription through an institution to be read.

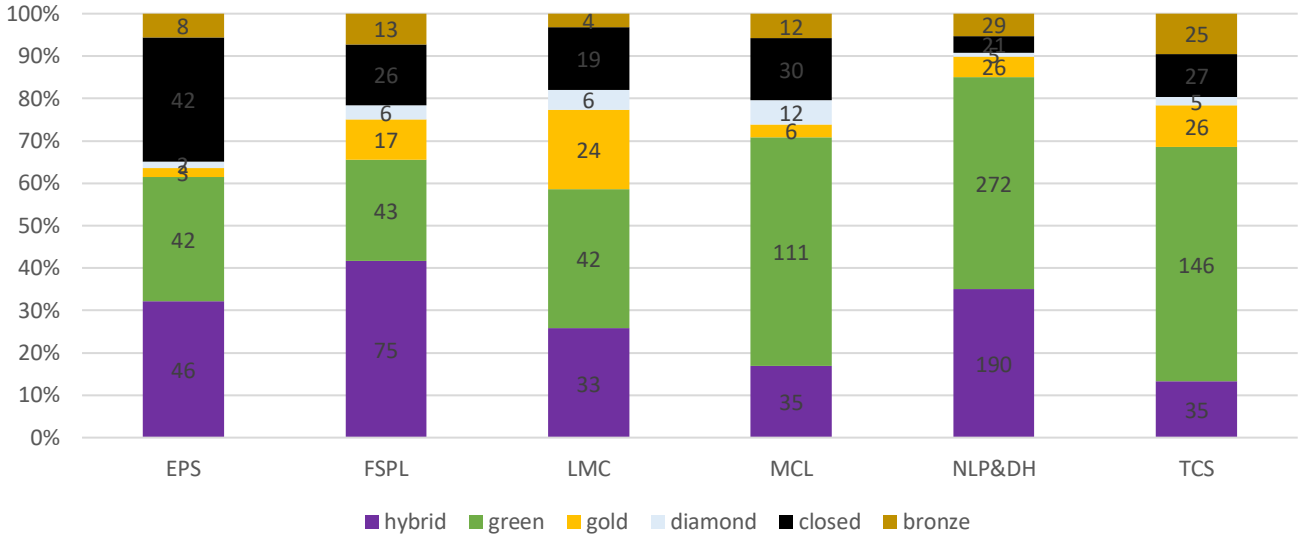
ILLC has published over 81% of its reportable output in one of the Open Access categories identified in OpenAlex. 43% percentage in green open access, 28% in hybrid access venues, 10% in gold open access, out of which 3% can be considered diamond. As shown in Figure 2, the ratio remains relatively consistent through the years, albeit with a slight decreasing trend in green open access publications.

Figure 2) Distribution of access type per publication year, ILLC reportable output, 2018-2023



When breaking down this information per research unit, we find wider variance, as shown in Figure 3. This can be largely attributed to differences in research culture and publication venues, which come to the fore in an interdisciplinary institute such as the ILLC. For instance, disciplines that produce a high volume of preprints have consequentially more green open access output. Of all the research units, Epistemology and Philosophy of Science (EPS) published over 34% of its output in closed access venues (closed and bronze access), the highest percentage of its overall output, and Natural Language Processing and Digital Humanities published the smallest proportion, just over 9%.

Figure 3) Distribution of access type per research unit, ILLC reportable output, 2018-2023



Impact and external collaboration

The ILLC is a significant player in its Core Area. Research published by ILLC authors frequently cites work from the most influential institutions in this field, including Stanford University, Oxford University, Cambridge University, Carnegie Mellon University, Washington University, and New York University. Conversely, these prestigious institutions often cite ILLC's work. As demonstrated in Tables 1 and 2, the top 20 institutions citing ILLC's work and being cited by ILLC remain relatively consistent and align with the most influential universities in the Core Area.

Table 1) Top 20 institutions citing ILLC works published in the period 2018-2023

Ranking	***Institution***	***Citations***
1	University of Amsterdam	1079
2	University of Edinburgh	256
3	Centre National de la Recherche Scientifique	132
4	Google	112
5	Tsinghua University	112
6	University of Cambridge	111
7	University of Oxford	111
8	Utrecht University	110
9	Radboud University Nijmegen	107
10	University of Washington	106
11	University College London	102
12	Stanford University	101
13	Carnegie Mellon University	97
14	New York University	88
15	University of Copenhagen	87
16	Vrije Universiteit Amsterdam	87
17	Ludwig-Maximilians-Universität München	86
18	Massachusetts Institute of Technology	85
19	Meta	80
20	University of California, Berkeley	78

Table 2) Top 20 most cited institutions in ILLC works, 2018-2023

Ranking	***Institution***	***Citations***
1	University of Amsterdam	1489
2	Stanford University	833
3	Massachusetts Institute of Technology	427
4	University of Edinburgh	404
5	University of Oxford	374
6	University of California, Berkeley	370
7	Google	364
8	Carnegie Mellon University	359
9	Cornell University	290
10	Princeton University	282
11	University College London	267
12	University of Washington	264
13	New York University	261
14	University of Cambridge	246
15	University of Toronto	244
16	Harvard University	242
17	Centre National de la Recherche Scientifique	225
18	University of California, San Diego	213
19	University of Groningen	186
20	Vrije Universiteit Amsterdam	185

Impact and internal collaboration

‘DORA-proof’ sophistication

An important recommendation, both from DORA and in the Leiden Manifesto, is to use a wide array of article-level metrics and to take into account differences between fields and in publication culture. Average SFNCI is derived from article-level metrics and normalised for differences between output types (e.g., journal articles, reviews, conference proceedings, book chapters) and between papers from different fields and publication years. However, they do not normalise for differences in publication culture within fields, e.g., as reflected in number of co-authors. That is why we included co-author count, and the number of affiliations involved in the OpenAlex export of publication sets to Excel, allowing closer inspection and further refinement of citation metrics.

Despite variations in the volume of output and impact across its research units, the ILLC consistently excels in terms of research impact within each field. NLP&DH produced the larger part of the output (over 38%) and received the larger share of citations at 61%. All the research units have a higher impact than expected in the research areas in which they publish, with at least 51% over the expectation, in the case of MCL, and as high as 602% higher than expected in the case of NLP&DH publications, followed by LMC and FSPL as close seconds. In terms of broader impact, research produced by LMC scores highly in its Altmetric attention score, Mendeley readers and number of article views, which indicates broader impact than its direct disciplinary embedding.

Table 3) Comparison ILLC Units

UNIT[1]	#P[2]	%P	C	%C	SFNCI	SD SFNCI	AUTH	SFNCI/ AUTH	INST	SFNCI/ INST	ALTMETRIC	MENDELEY	VIEWS
EPS	143	9,6%	991	5,3%	4,70	9,15	2,64	2,10	1,80	3,03	3,54	12,04	11,91
FSPL	180	12,1%	2273	12,2%	5,14	13,01	2,48	2,33	1,57	3,55	7,26	12,52	17,69
LMC	123	8,2%	1083	5,8%	5,56	14,76	3,39	1,60	2,02	2,78	15,21	21,25	20,00
MCL	206	13,8%	710	3,8%	1,51	2,50	3,07	0,55	1,75	0,78	1,53	5,60	8,99
NLP&DH	575	38,6%	11505	61,6%	6,02	28,12	4,25	1,81	1,62	4,08	10,53	64,26	27,48
TCS	264	17,7%	2110	11,3%	2,25	5,87	3,66	0,55	2,13	0,86	4,78	14,60	10,82
ILLC_ALL	1519		18298		4,33	18,98	3,48	1,50	1,82	2,73	6,66	28,00	14,73

[1] EPS: Epistemology & Philosophy of Science; FSPL: Formal Semantics & Philosophical Logic; LMC: Language & Music Cognition; MCL: Mathematical & Computational Logic (MCL); NLP&DH: Natural Language Processing & Digital Humanities; TCS: Theoretical Computer Science.
[2] #P: number of publications; %P percentage of publications; #C: number of citations; %C percentage of citations; SFNCI: subfield normalized citation impact; SD SFWCI: Standard deviation of subfield normalized citation impact; #AUTH: number of co-Authors; SFNCI/AUT: subfield normalized citation impact divided by number of authors; #INST: number of co-institutions; SFNCI/INST: subfield normalized citation impact divided by number of institutions; ALTMET: altmetric attention score; MENDELEY: total number of mendeley readers; VIEWS: Views Counts are generated from usage data in Scopus; See Appendix 1 for all metric definitions

Table 4) Internal co-authorship for ILLC research units, reportable output 2018-2023

	NLP&DH	TCS	MCL	FSPL	EPS	LMC
NLP&DH	66	0	0	21	3	9
TCS	0	30	7	1	1	0
MCL	0	7	24	8	16	0
FSPL	21	1	8	28	10	14
EPS	3	1	16	10	17	0
LMC	9	0	0	14	0	4

Table 4 represents co-authorship of publications between ILLC authors. Every unique authorship pairing of two distinct units (including a unit with itself) is counted as one. In broad lines, authors collaborate most frequently with other members of their own unit, with the only exception of Language and Music Cognition (LMC). Authors belonging to the unit Formal Semantics and Philosophical Logic (FSPL) collaborate most frequently institute wide, and Theoretical Computer Science (TCS) authors co-author papers with other units the least.

Figure 4) relative contribution of UvA in the ILLC Core Area (main topics, 2000-2023)

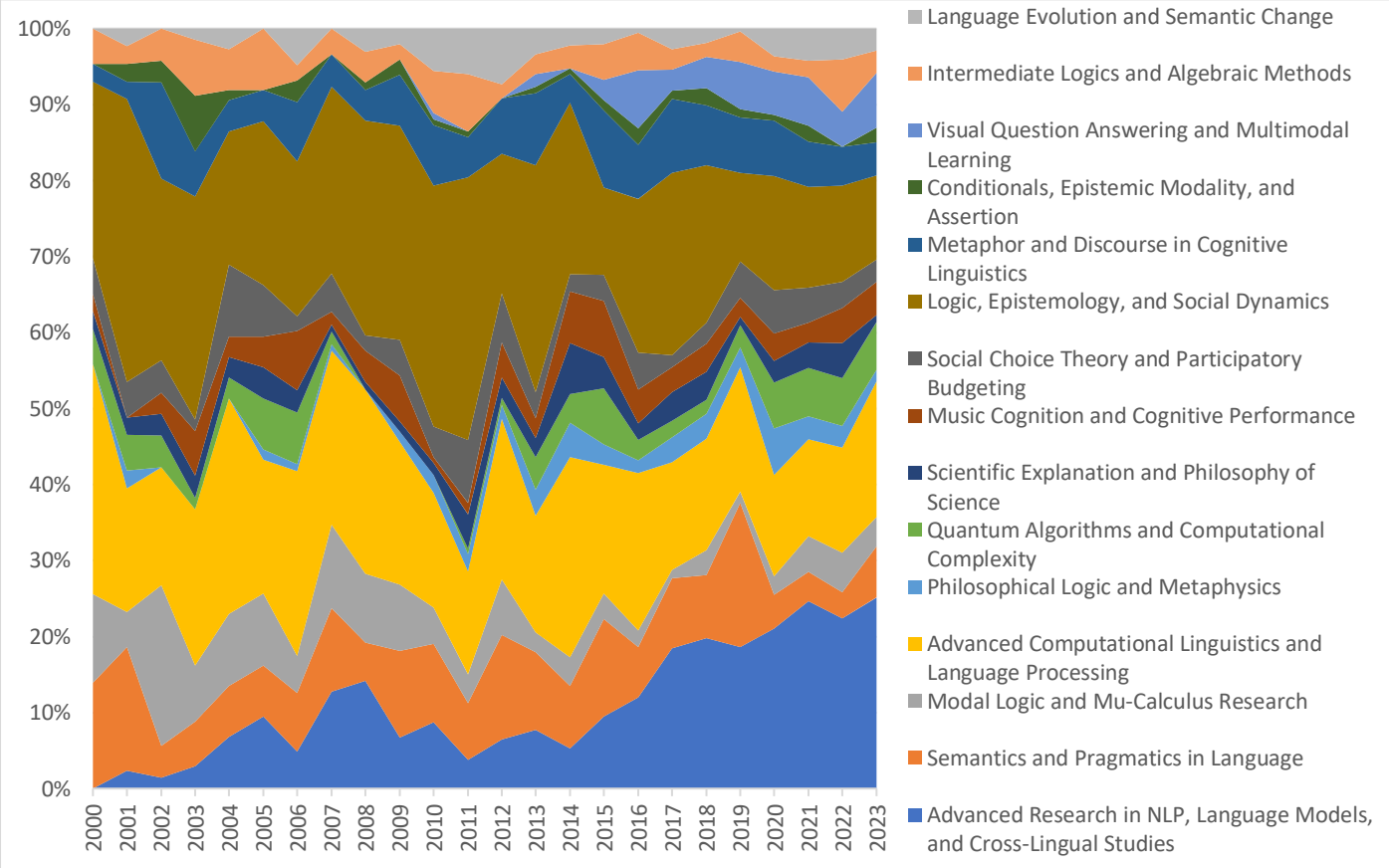
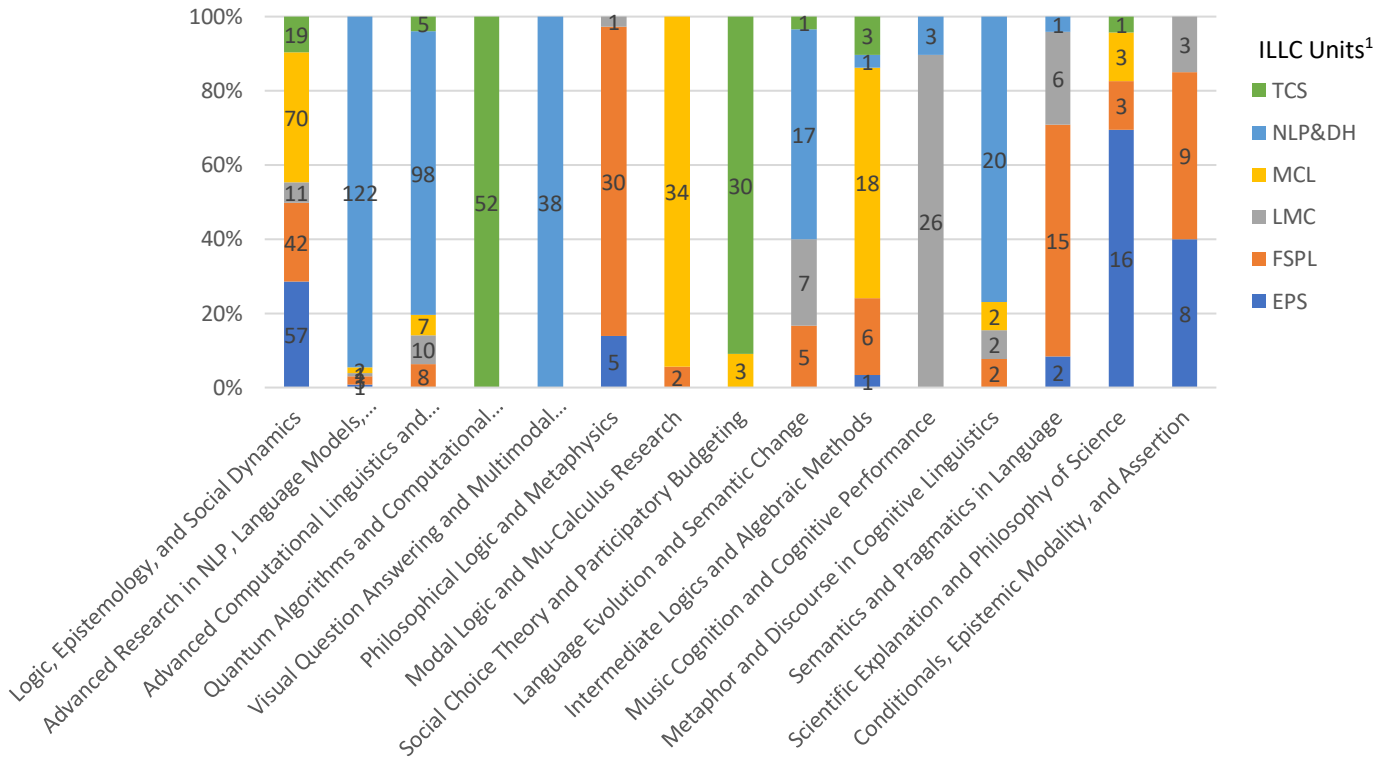


Figure 4 illustrates the University of Amsterdam's (UvA) relative contribution to the ILLC Core Area from 2000 to 2023. The topics “Visual Question Answering and Multimodal Learning” and “Advanced Research in NLP, Language Models, and Cross-Lingual Studies” have shown notable growth during this period. Topics such as “Advanced Computational Linguistics and Language Processing,” “Logic, Epistemology, and Social Dynamics,” “Semantics and Pragmatics in Language,” “Metaphor and Discourse in Cognitive Linguistics,” “Music Cognition and Cognitive Performance,” and “Intermediate Logics and Algebraic Methods” have maintained stable development and a consistent presence. Conversely, topics like “Conditionals, Epistemic Modality, and Assertion,” “Quantum Algorithms and Computational Complexity,” and “Modal Logic and Mu-Calculus Research” have exhibited more intermittent activity. Over the past six years, UvA's involvement in the various topics within the ILLC Core Area has remained relatively unchanged.

Figure 5) Relative distribution ILLC units publications over 15 Core Area topics



[1] EPS: Epistemology & Philosophy of Science; FSPL: Formal Semantics & Philosophical Logic; LMC: Language & Music Cognition; MCL: Mathematical & Computational Logic (MCL); NLP&DH: Natural Language Processing & Digital Humanities; TCS: Theoretical Computer Science.

Figure 5 illustrates the presence of the six ILLC units within the 15 Core Area topics. Each unit exhibits a dominant presence in at least one Core Area topic. The largest topic, “Logic, Epistemology, and Social Dynamics,” encompasses contributions from all units except NLP&DH. Notably, the NLP&DH unit dominates five topics. Two units are dominant in a single topic each: Language & Music Cognition (LMC) leads in “Music Cognition and Cognitive Performance,” while Epistemology & Philosophy of Science (EPS) is prominent in “Scientific Explanation and Philosophy of Science.”

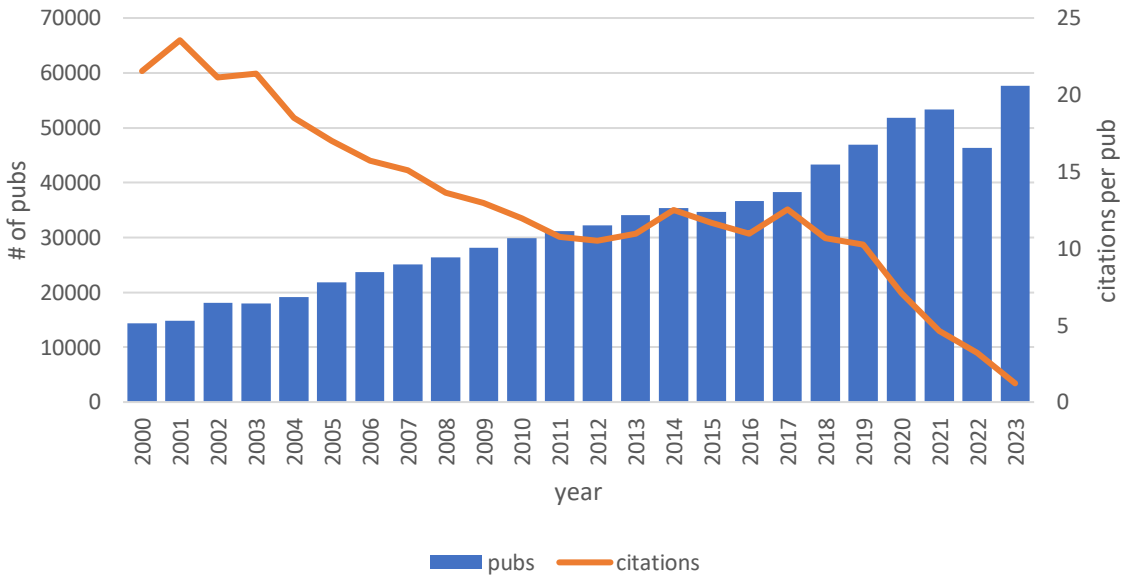
Global trends for the ILLC Core Area

We analysed the ILLC Core Area for the period 2000-2023. For over 200,000 publications, we assessed impact, collaboration, open access status, and trends over time. Subsequently, we compared the 12 most productive institutions worldwide in terms of impact, collaboration, open access status, and presence and development over time.

Global trends in the Core Area, 2000-2023

Figure 6 illustrates global activity in the ILLC Core Area, showcasing the number of publications and citations per publication from 2000 to 2023. There is a nearly linear, constant rise in output, with the exception of 2022, which shows a notable anomaly with significantly lower production compared to 2021 and 2023. The reason for this anomaly has yet to be determined. The number of citations per publication increases with the age of the publications, which is an expected trend; older publications have had more time to accumulate citations.

Figure 6) Global activity in ILLC Core Area # pubs and citations per pub. 2000-2023



While Figure 5 illustrates the relative contribution of UvA publications within the ILLC Core Area, Figure 7 depicts the relative contribution of all worldwide publications in the ILLC Core Area topics from 2000 to 2023. Due to the significantly larger number of publications, Figure 7 exhibits a smoother, less spiky pattern. The overall trends remain similar: the topic “Visual Question Answering and Multimodal Learning” has emerged around 2015, growing from less than one percent to over ten percent in 2023. Meanwhile, “Advanced Computational Linguistics and Language Processing” has shown a steep rise, increasing from approximately five percent in 2015 to more than twenty percent in 2023. Additionally, the topic “Quantum Algorithms and Computational Complexity” has demonstrated notable growth from around 2019 to the present.

Figure 7) relative contribution of all (worldwide) publications for ILLC Core Area topics (2000-2023)

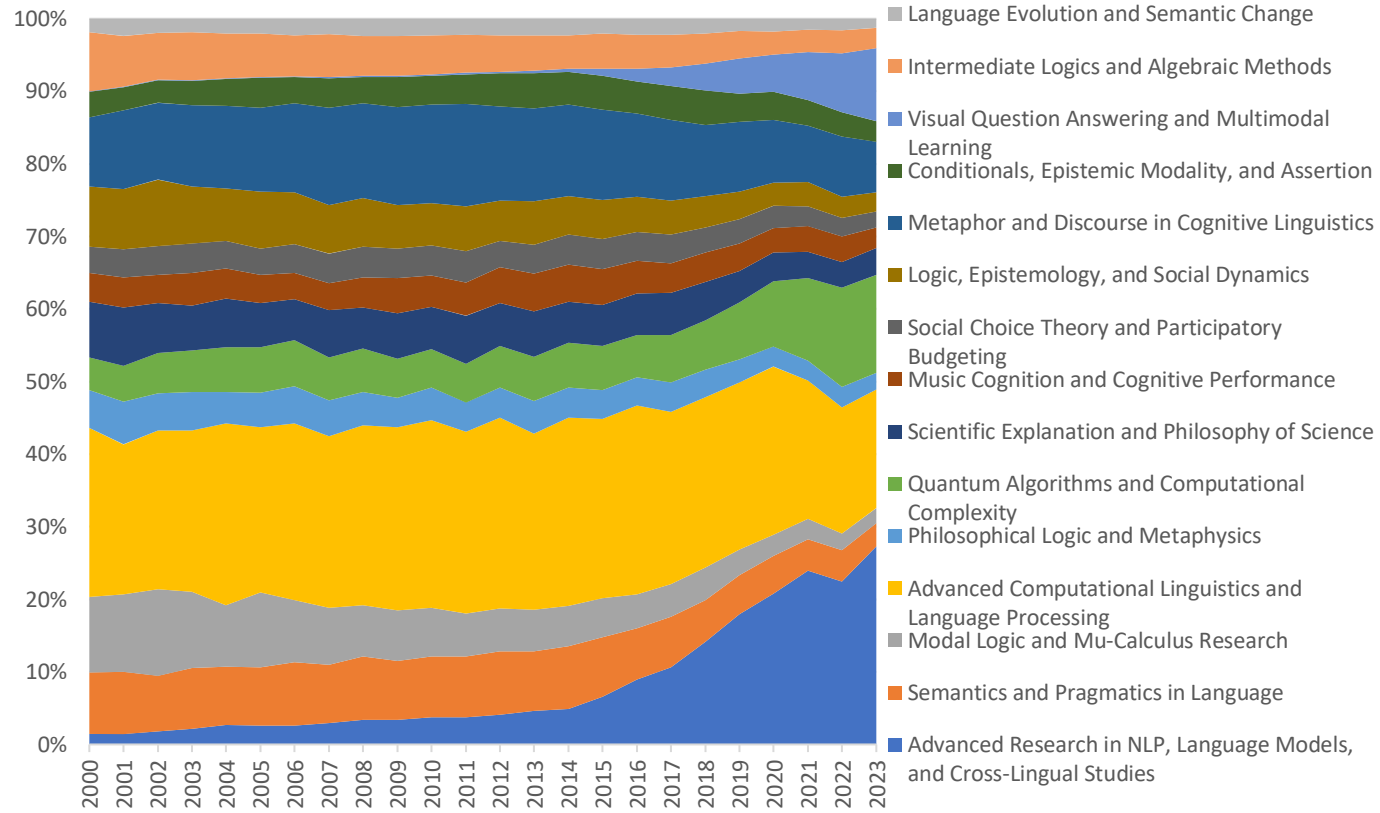


Figure 8) International/ Global Open Access status ILLC Core Area 2000-2023

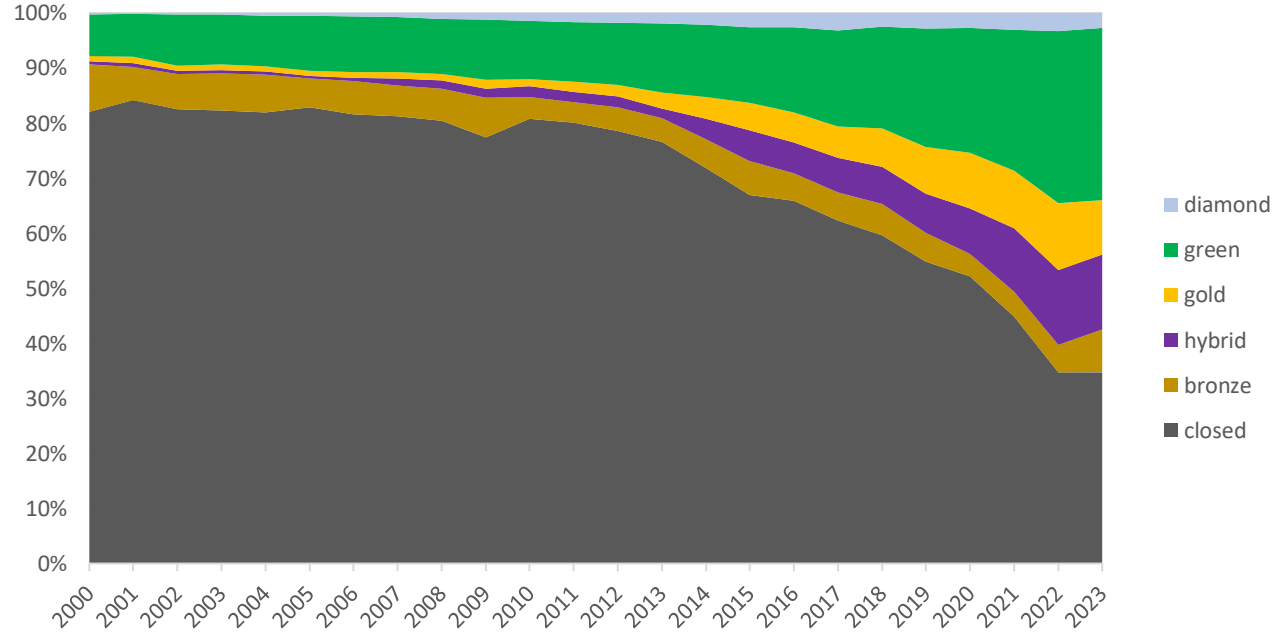


Figure 8 depicts the Open Access status of all OpenAlex publications within the ILLC Core Area from 2000 to 2023. Notably, Figure 2 highlights the Open Access distribution specifically for ILLC, revealing a significantly lower proportion of closed access publications compared to the overall Core Area. Bronze[1]

publications are considered closed access, which meant that in 2000, ninety percent of all publications were closed. This has dramatically decreased to around forty percent by 2023. Consequently, all other forms of access have benefited from this increase in openness, with green open access publications showing the most significant growth.

[1] Bronze articles are free to read on the publisher's website, without a license that grants any other rights. There may be a delay between publication and availability to read, and often articles can be removed unilaterally by the publisher.

Bibliometric benchmark of Top-12 universities in the Core Area

Table 5) Top 12 most productive Universities in ILLC Core Area 2018-2023 and 13 metrics

UNIVERSITY[1]	#P	%P	#C	%C	SFNCI	SD SFNCI	#-AUTH	SFNCI/ AUTH	NR INST	SFNCI/ INST	ALTMET	MEN DELEY	VIEWS
CARNEGIE	1917	11,2%	36643	12,3%	3,95	12,55	5,03	1,02	2,62	2,01	6,75	57,05	12,18
TSINGHUA	1834	10,7%	29218	9,8%	3,62	9,68	5,59	0,72	2,85	1,54	4,49	36,00	17,74
PEKING	1735	10,2%	24833	8,3%	2,68	5,80	4,93	0,58	2,64	1,17	4,48	24,52	13,43
UOFCAS	1459	8,5%	19063	6,4%	2,70	8,99	5,02	0,56	3,53	0,80	3,09	27,23	12,76
OXFORD	1441	8,4%	18352	6,1%	6,25	26,06	3,49	2,32	2,67	2,56	13,12	27,16	12,57
EDINBURGH	1413	8,3%	19526	6,5%	5,42	15,67	4,40	2,17	2,86	2,88	7,07	39,88	13,50
CAMBRIDGE	1394	8,2%	15621	5,2%	6,17	21,08	4,27	2,32	2,85	2,69	10,35	32,58	12,64
AMSTERDAM	1352	7,9%	15821	5,3%	4,18	13,59	3,68	1,38	2,74	1,92	4,56	26,98	12,26
STANFORD	1310	7,7%	29963	10,0%	6,62	17,45	4,49	2,09	2,58	3,59	23,77	58,96	14,35
WASHINGTON	1104	6,5%	48717	16,3%	8,71	33,81	5,35	1,82	2,97	3,50	23,85	118,02	12,28
NYU	1084	6,3%	23858	8,0%	6,48	21,35	4,25	2,68	2,76	3,49	11,78	40,88	13,98
UNICOLLON	1047	6,1%	17033	5,7%	7,76	30,15	4,31	1,78	3,15	2,33	16,13	36,53	14,91
	17090		298648										
GLOBAL CA	281002		1745714		2,76	24,17	2,99	1,18	1,05	1,83			
ILLC_ALL	1519		18298		4,33	18,98	3,48	1,50	1,82	2,73	6,66	28,00	14,73
ILLC_CA	825		8877		3,77	11,82	3,14	1,50	1,76	2,60	3,92	27,95	11,57
ILLC_OTHER	694		9421		5,02	25,24	3,89	1,49	1,90	2,89	10,01	28,06	18,37

[1] Carnegie Mellon University (CARNEGIE); Tsinghua University (TSINGHUA); Peking University (PEKING); University of Chinese Academy of Sciences (UOFCAS); University of Oxford (OXFORD); University of Edinburgh (EDINBURGH); University of Cambridge (CAMBRIDGE); University of Amsterdam (AMSTERDAM); Stanford University (STANFORD); University of Washington (WASHINGTON); New York University (NYU); University College London (UNICOLLON).

The ILLC Core Area has produced an impressive 280,000 publications in OpenAlex from 2018 to 2023. Table 5 highlights the top 12 most productive institutions, excluding joint research institutes and corporate R&D departments (Google, Microsoft) from the analysis. Leading with a slight margin, Carnegie Mellon University (CARNEGIE) emerges as the most productive institution. The University of Amsterdam (AMSTERDAM) holds the eighth position. For a comprehensive list of institution abbreviations, please refer to the footnote of Table 5.

The UvA dataset includes (1352-825=) 527 papers that are not co-authored by ILLC, which are from other UvA institutes. For this reason, the bottom three rows of Table 5 show the metrics for all ILLC papers both inside and outside its Core Area, as well as for its total publication set.

ILLC's 825 Core Area papers would rank 25th globally in terms of publication count among universities. However, it is inevitable that most universities listed will also include papers from units outside of their

'Logic, Language and Computation' departments. Therefore, using an international benchmark at the university level rather than the department level may be the fairest approach.
It is noteworthy that non-Core Area ILLC publications exhibit a higher SFNCI compared to Core Area ILLC publications. A closer examination of the individual publication lists reveals that this discrepancy is due to a few exceptionally high-impact publications in the non-Core Area. This is further evidenced by the SFNCI standard deviation, which is 12 for the Core Area and 25 for the non-Core Area.

Table 6) Metrics ranking of top 12 most productive Universities in ILLC Core Area 2018-2023

UNIVERSITY	#P	%P	#C	%C	SFNCI	SD SFNCI	#-AUTH	SFNCI/ AUTH	NR INST	SFNCI/ INST	ALTMET	MEN DELEY	VIEWS
CARNEGIE	1	1	2	2	9	9	4	3	9	11	8	3	12
TSINGHUA	2	2	4	4	10	10	3	1	10	6	10	7	1
PEKING	3	3	5	5	12	12	1	5	11	10	11	12	6
UOFCAS	4	4	8	8	11	11	2	4	12	1	12	9	7
OXFORD	5	5	9	9	5	3	10	12	2	9	6	10	9
EDINBURGH	6	6	7	7	7	7	6	7	4	4	4	5	5
CAMBRIDGE	7	7	12	12	6	5	8	9	3	5	5	8	8
AMSTERDAM	8	8	11	11	8	8	5	11	8	8	9	11	11
STANFORD	9	9	3	3	3	6	7	6	5	12	1	2	3
WASHINGTON	10	10	1	1	1	1	12	2	6	3	2	1	10
NYU	11	11	6	6	4	4	9	10	1	7	3	4	4
UNICOLLON	12	12	10	10	2	2	11	8	7	2	7	6	2

Table 7) Co-authorship for top 12 universities in ILLC Core Area 2018-2023

	Peking	Cambridge	Oxford	UofCAS	UniColLon	New York	Carnegie	Amsterdam	Stanford	Edinburgh	Tsinghua
Washington	9	5	3	1	12	24	73	16	29	11	2
Peking		6	17	37	2	5	8	3	13	5	89
Cambridge			37	1	36	5	9	15	6	33	11
Oxford				1	33	3	13	62	9	19	13
UofCAS					2	1	1	4	2		39
UniColLon						6	3	18	5	15	1
New York							29	9	18	10	1
Carnegie								5	29	17	20
Amsterdam									15	78	18
Stanford										13	12
Edinburgh											9

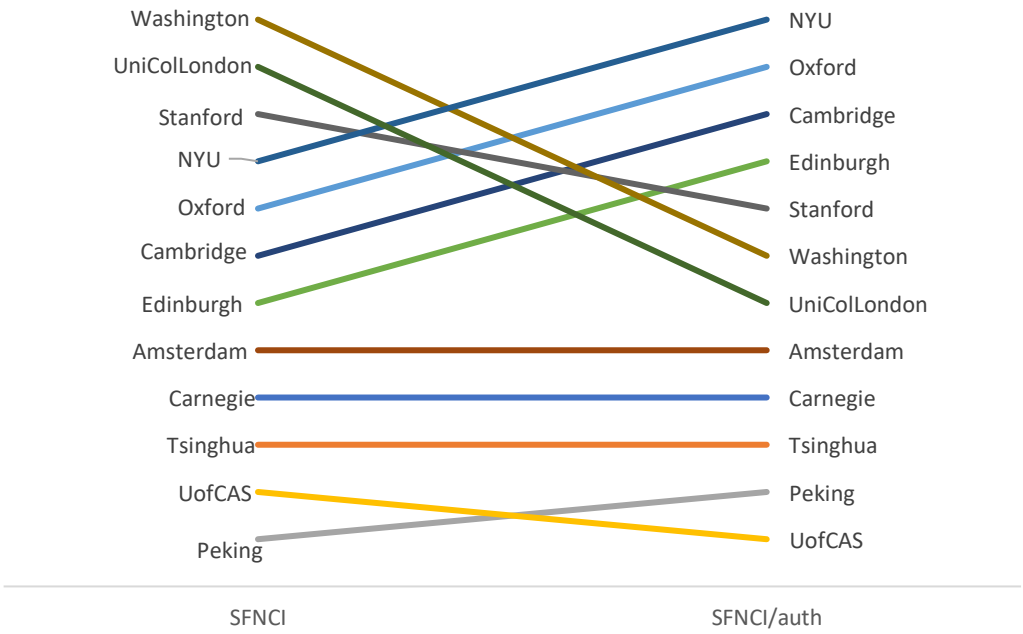
Table 7 highlights co-authorship in the ILLC Core Area from 2018 to 2023. A strong national collaboration trend is evident, with Chinese, British, and American universities displaying a high number of national joint publications. Notably, there is a significant international partnership between the University of Edinburgh and the University of Amsterdam, with 78 joint publications, and between the University of Oxford and the University of Amsterdam, with 62 joint publications.

Table 8 presents the impact of joint publications in the ILLC Core Area for the top 15 institutions, focusing on the average impact (SFNCI) of substantial co-authorships (ten or more joint publications). Collaborations with other top 15 ILLC Core Area institutions positively influence the impact for the University of Amsterdam (UvA). Notably, partnerships with the University of Washington (SFNCI = 19.7), University College London (SFNCI = 20.2), University of Cambridge (SFNCI = 6.0), and University of Edinburgh (SFNCI = 7.1) have shown significant positive effects. Additionally, British and American universities demonstrate a clear strengthening effect on impact through co-authorship with national partners.

Table 8) SubField Normalized Citation Impact for top 12 universities in ILLC Core Area 2018-2023

	Peking	*Cambridge*	*Oxford*	*UofCAS*	*UniColLon*	*New York*	*Carnegie*	*Amsterdam*	*Stanford*	*Edinburgh*	*Tsinghua*
Washington					4,64	26,61	7,14	19,73	9,32	21,09	
Peking			10,91	2,56							3,83
Cambridge			12,96		24,02			6,01		8,90	1,48
Oxford					32,23		2,51	3,59		14,46	6,14
UofCAS											5,24
UniColLon								20,25		2,19	
New York							7,82		5,17	1,59	
Carnegie									7,14	4,86	3,74
Amsterdam									2,82	7,14	3,37
Stanford										10,44	1,31
Edinburgh											

Figure 9) SFNCI and SFNCI/auth ranking compared for top 12 institutions ILLC Core Area 2018-2023



The University of Amsterdam (UvA) maintains a stable position and high ranking in the ILLC Core Area across all metrics, as shown in Table 6 and Figure 9. Even when normalized for differences in publication culture within the field, UvA holds its ground. Specifically, for SFNCI and SFNCI/auth metrics, UvA ranks eighth among the top 12 most productive institutes globally. Figure 9 illustrates the dynamic ranking shifts among the top seven most impactful institutes within the ILLC Core Area. These fluctuations highlight how differences in the field itself, including variations in the number of co-authors and the number of participating institutes, can influence rankings.

Conclusions

This international benchmark clearly shows that the Institute for Logic, Language, and Computation has been one of the leading institutions in its Core Area during the period 2018-2023. This multidisciplinary institute overperforms in all its different focus areas and ranks high among the top universities in its Core Area. Likewise, the ILLC has strong international collaborative ties with these institutions, receiving citations from and citing work from these top institutions. From an internal perspective, the fact that the top 15 topics of the Core Area represent 56% of the total output of the institute, indicates the coherence of its scientific output, since OpenAlex topic classification is primarily based on citation patterns, and other text-based features, like the title and abstract are only secondarily considered. In other words, a significant majority of ILLC publications (its Core Area) are part of a densely knit citation network.