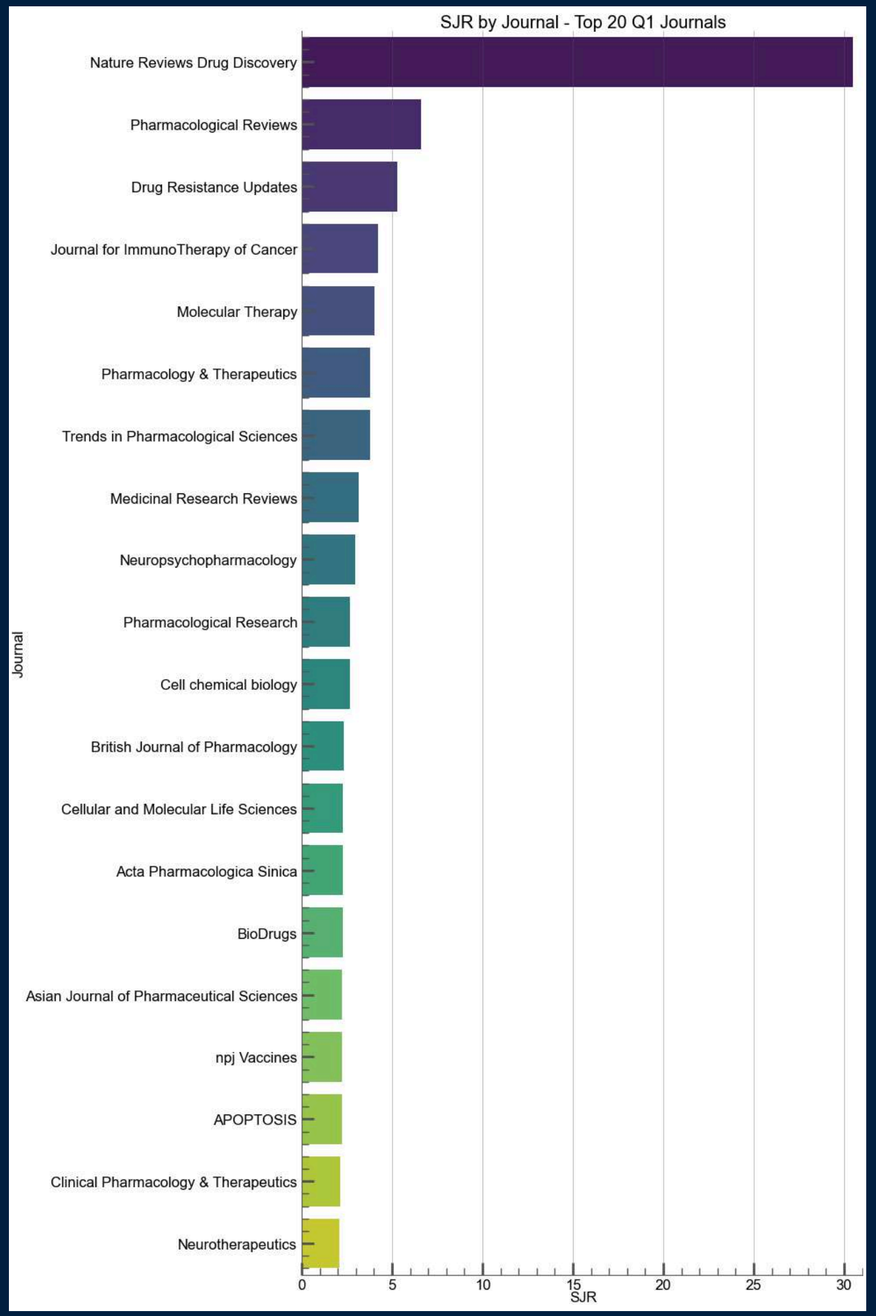
BJP Among Q1 Journals

A Data-Driven Analysis of Impact and Performance

Noah Roussel - Under the supervision of Csaba Kiss and Dr. Roland Molontay

HSDSLab
Human & Social Data Science



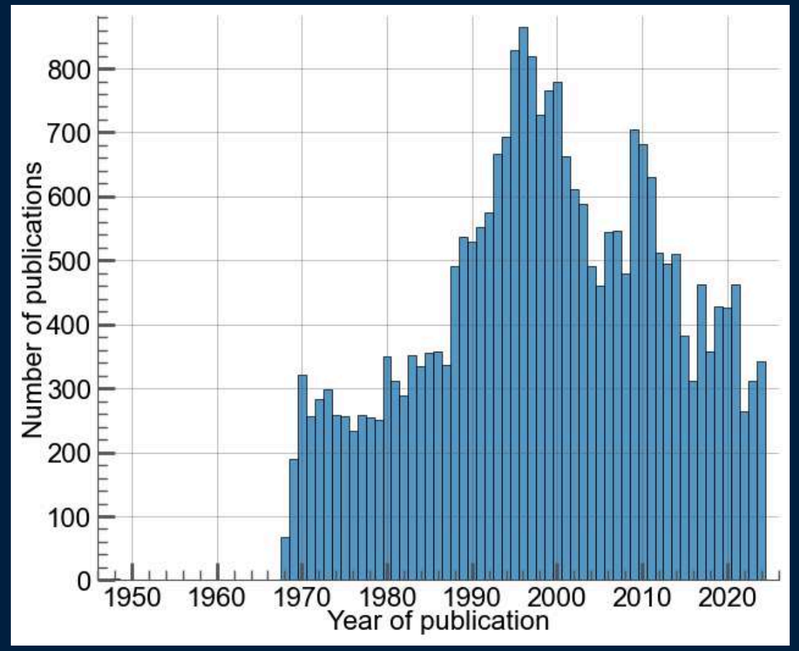
Starting Point : BJP's SJR Score

- Evaluation based on Scopus SJR Indicator for Q1 pharmacology journals (81 journals).
- BJP shows a solid ranking (12th place)
  (Plot only showing the top 20 journals for better visualization)
- Next step : Understanding the factors behind this performance

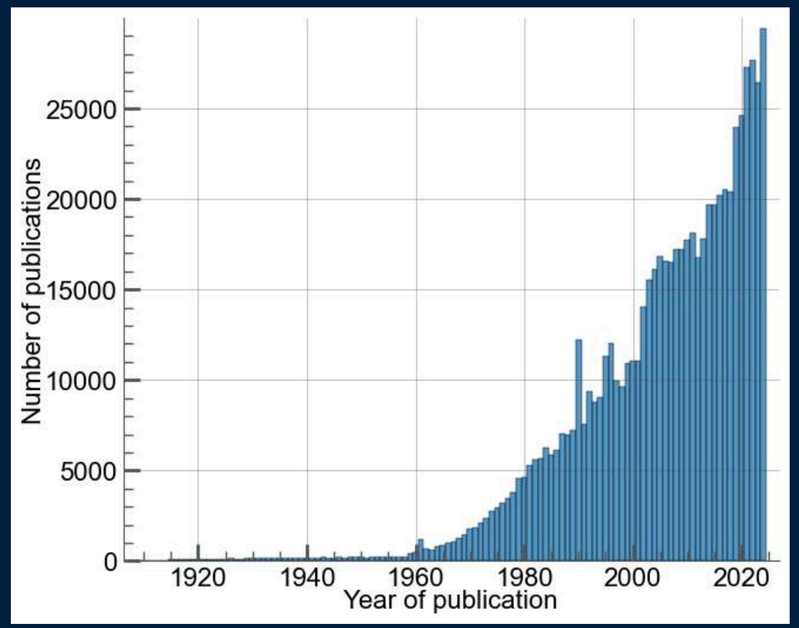
HSDSLab
Human & Social Data Science

Data

- Data collected from OpenAlex, an open-source database of scientific publications.
- 689 301 works from Q1 journals (with BJP) from 1909 to 2024
- 26 115 works from BJP Journal between 1948 and 2024
- Starting the analysis from 1950



BJP Journal Evolution of Publications



Q1 Journals Evolution of Publications

HSDSLab
Human & Social Data Science

Comparative Analysis Agenda

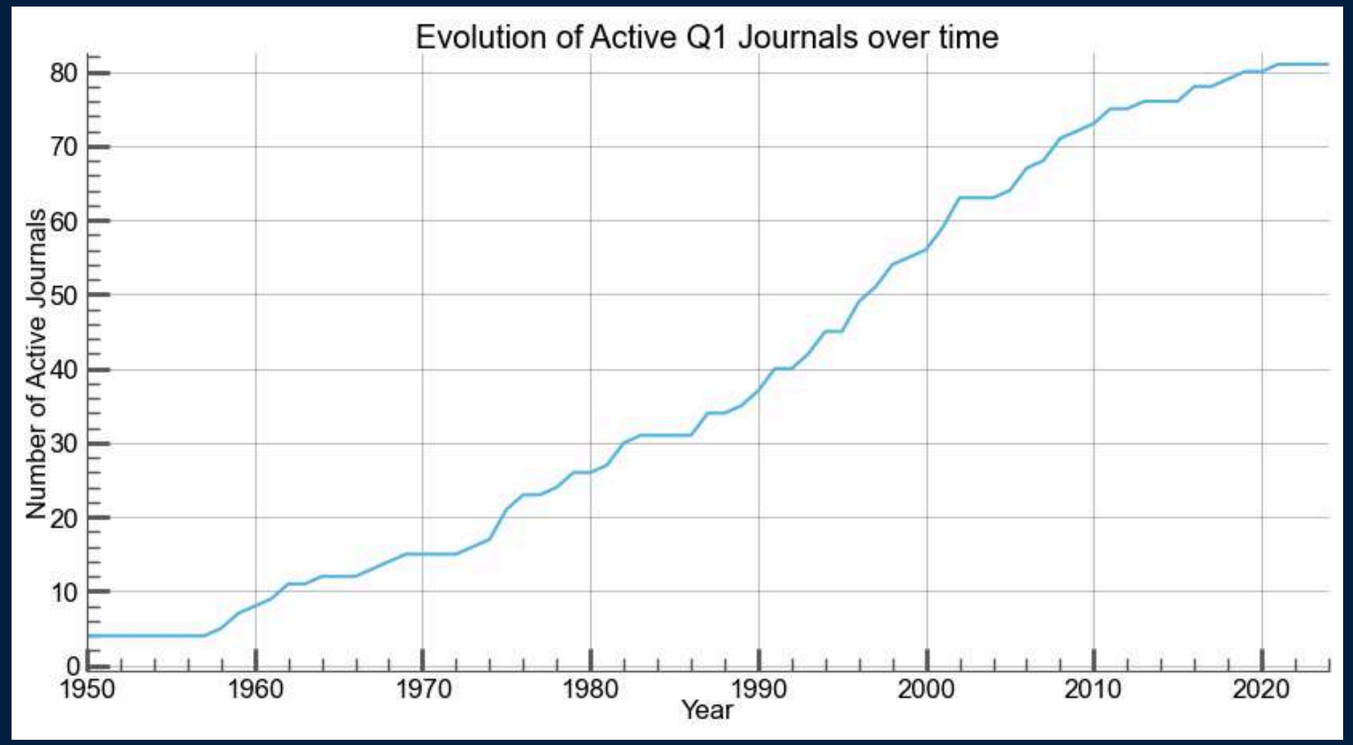
1. General Metrics: Key Indicators of Journal Impact
2. Author Overview: Contributions and Gender Distribution
3. Countries & institutions: Impact, rankings & global collaborations
4. Topics: Main research areas & modelling
5. Modelling: Key trends and performance metrics

HSDSLab
Human & Social Data Science

General Metrics

General Metrics

First-Year Work Analysis - Q1 Journals



Journal	First Year
Journal of Pharmacology and Experimental Therapeutics	1909
Cellular and Molecular Life Sciences	1945
British Journal of Pharmacology	1948
Pharmacological Reviews	1949
Biochemical Pharmacology	1958
Antiviral Research	1959
Psychopharmacology	1959
Clinical Pharmacology & Therapeutics	1960
PharmacoEconomics	1961
Neuropharmacology	1962
Antimicrobial Agents and Chemotherapy	1962
Molecular Pharmacology	1964
European Journal of Pharmacology	1967
Clinical Trials	1968
Current Neuropharmacology	1969
Drug Metabolism and Disposition	1973
British Journal of Clinical Pharmacology	1974
Pain	1975
Drug and Alcohol Dependence	1975
Inflammation Research	1975
Journal of Antimicrobial Chemotherapy	1975
Pharmacology & Therapeutics	1976
Clinical Pharmacokinetics	1976
Neuropsychopharmacology	1978
Trends in Pharmacological Sciences	1979
Journal of Ethnopharmacology	1979
Medicinal Research Reviews	1981

HSDSLab
Human & Social Data Science

General Metrics

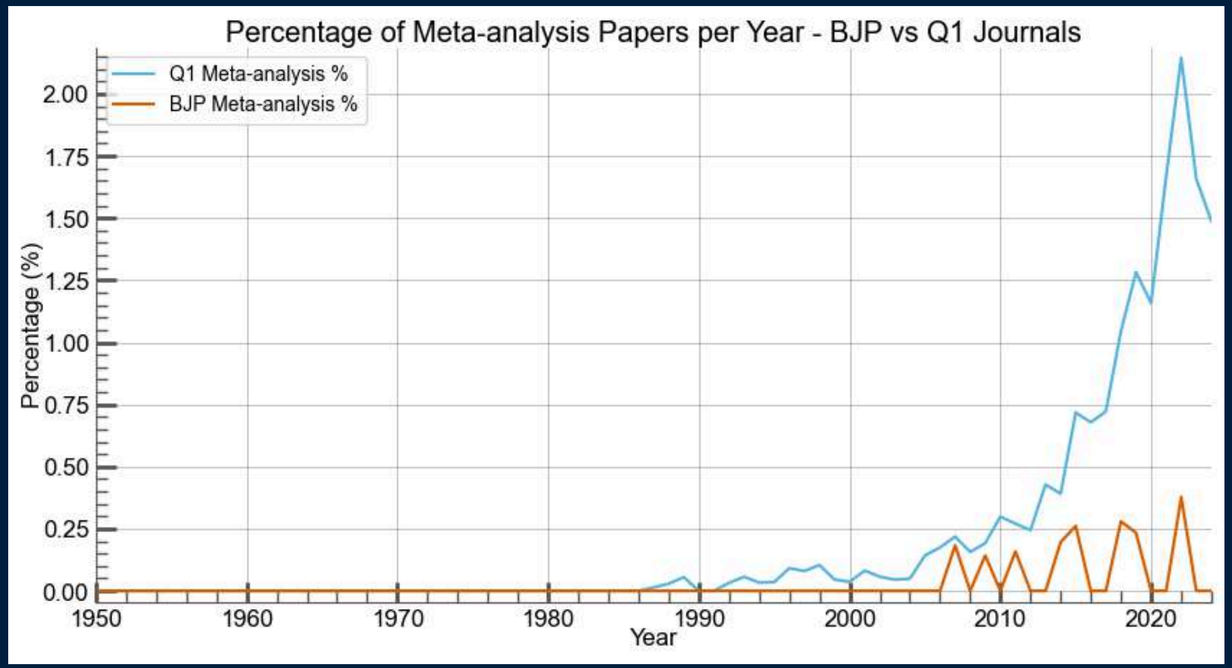
First-Year Work Analysis - Q1 Journals

Journal	First Year
Progress in Neuro-Psychopharmacology and Biological Psychiatry	1982
Biomedicine & Pharmacotherapy	1982
Drug Safety	1982
Investigational New Drugs	1983
Journal of Psychopharmacology	1987
Phytotherapy Research	1987
European Journal of Medicinal Chemistry	1987
Pharmacological Research	1989
European Neuropsychopharmacology	1990
Bioconjugate Chemistry	1990
Inflammopharmacology	1991
Journal of Immunotherapy	1991
Vaccines	1991
Expert Opinion on Therapeutic Patents	1993
Biomolecules & Therapeutics	1993
APOPTOSIS	1994
Phytomedicine	1994
Expert Opinion on Investigational Drugs	1994
Neurotherapeutics	1996
Drug Discovery Today	1996
Addiction Biology	1996
Environmental Toxicology and Pharmacology	1996
BioDrugs	1997
Expert Opinion on Emerging Drugs	1997
Drug Resistance Updates	1998
The International Journal of Neuropsychopharmacology	1998
Expert Opinion on Therapeutic Targets	1998
Acta Pharmacologica Sinica	1999
Molecular Therapy	2000
Current Opinion in Pharmacology	2001
International Immunopharmacology	2001
Expert Opinion on Biological Therapy	2001
Nature Reviews Drug Discovery	2002
Expert Review of Vaccines	2002
Cancer Biology & Therapy	2002
Pharmaceutical Statistics	2002
EXCLI journal	2005
Molecular Diagnosis & Therapy	2006
Journal of Neuroimmune Pharmacology	2006
Chinese Medicine	2006
Drug Design Development and Therapy	2007
CNS Neuroscience & Therapeutics	2008
Therapeutic Advances in Neurological Disorders	2008
Pharmaceutical Medicine	2008
Asian Journal of Pharmaceutical Sciences	2009
Frontiers in Pharmacology	2010
Journal of Xenobiotics	2011
Natural Products and Bioprospecting	2011
Journal for ImmunoTherapy of Cancer	2013
Cell chemical biology	2016
npj Vaccines	2016
ACS Pharmacology & Translational Science	2018
Medicine in Drug Discovery	2019
NEJM Evidence	2021

HSDSLab
Human & Social Data Science

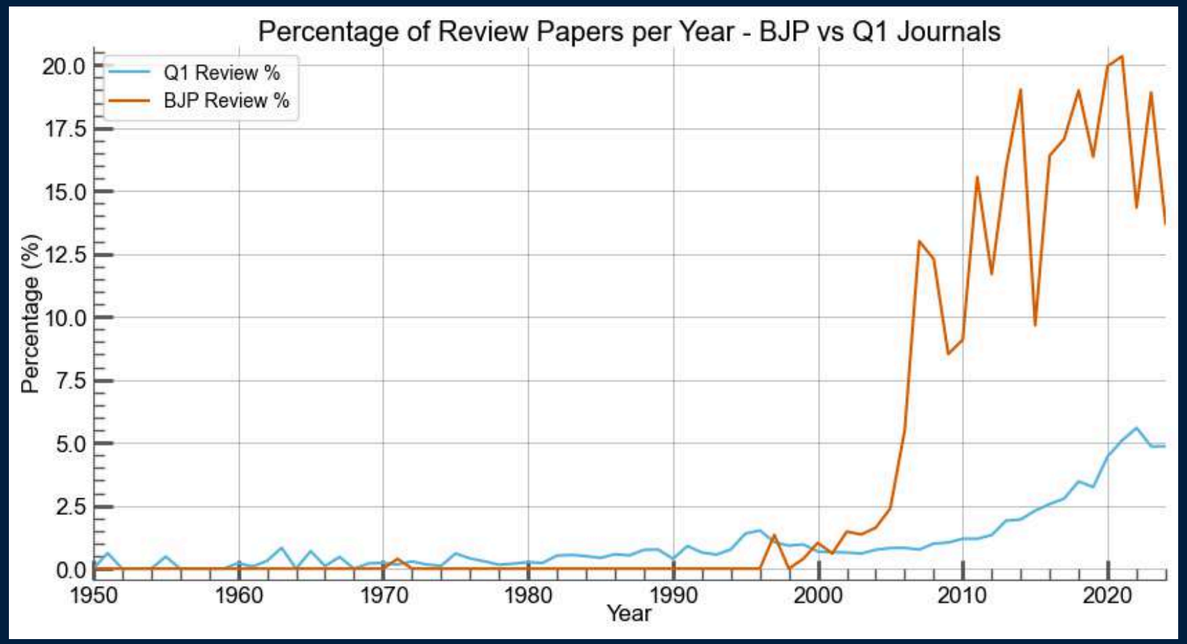
General Metrics

Review and Meta-analysis %



More meta-analyses for Q1 Journals

More reviews for the BJP



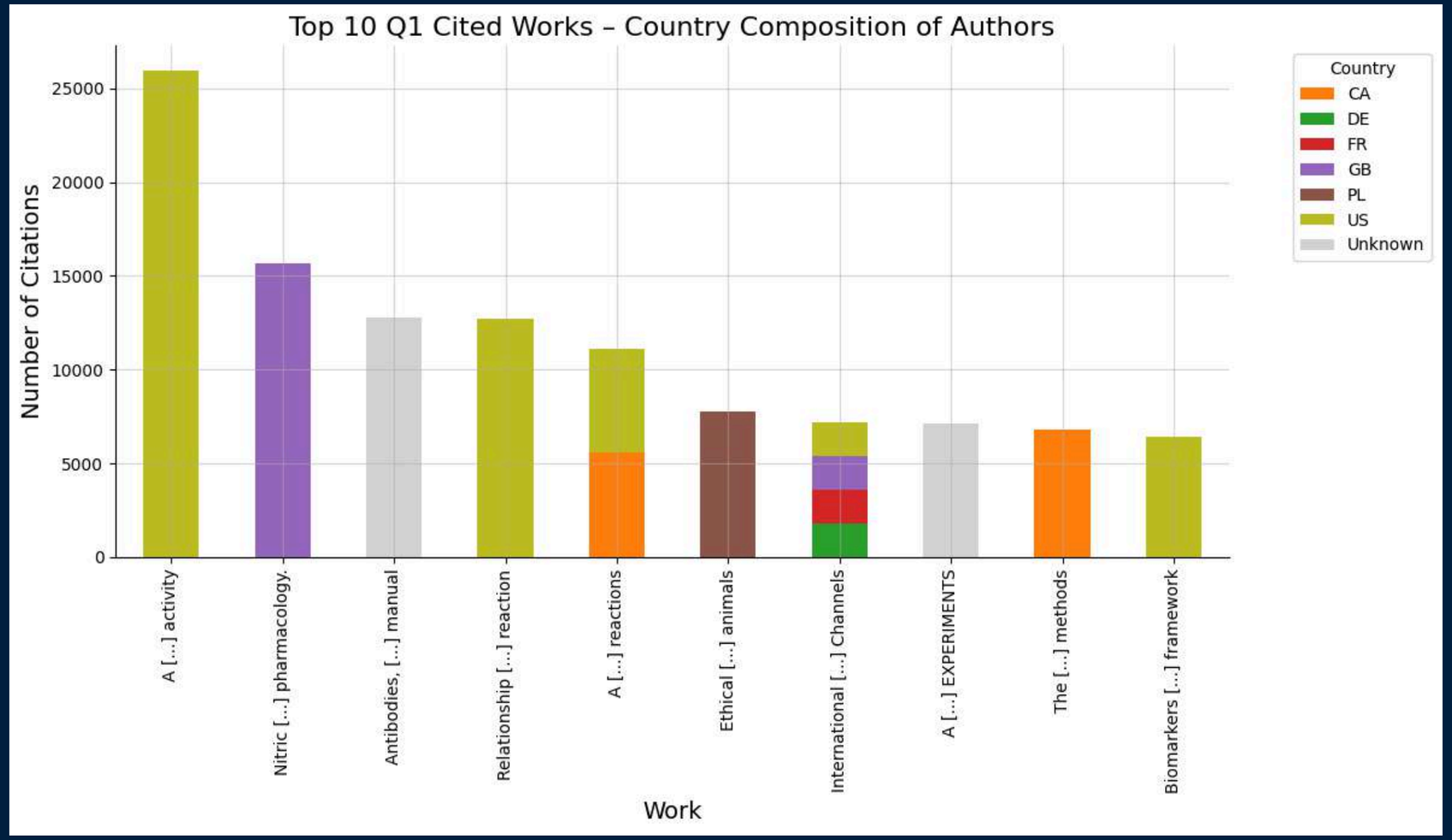
HSDSLab
Human & Social Data Science

General Metrics

« *A new and rapid colorimetric determination of acetylcholinesterase activity* », 1961 is the most cited work (25983).

Journal : *Biochemical Pharmacology*

Top 10 Works - Q1 Journals



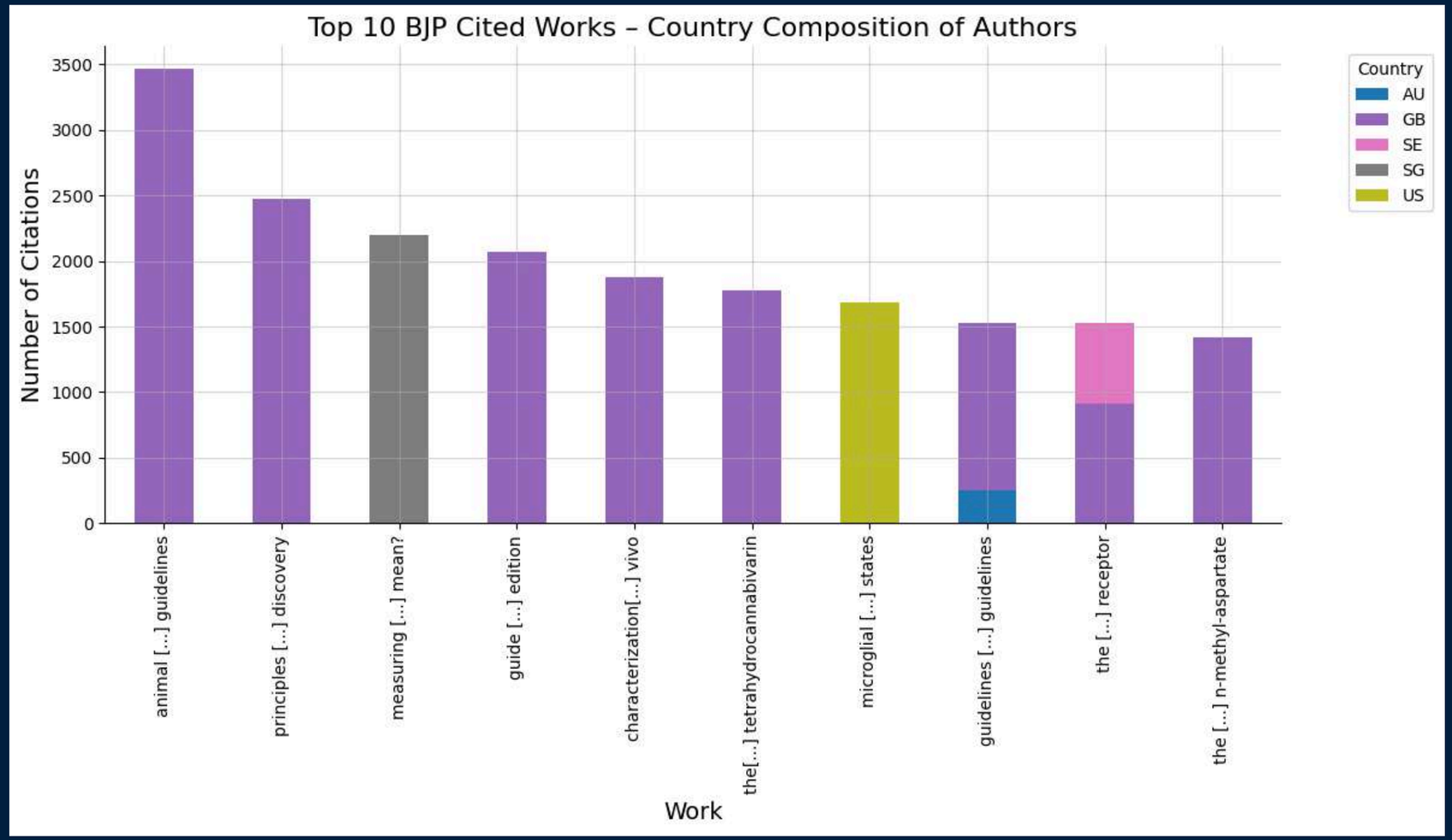
HSDSLab
Human & Social Data Science

General Metrics

« Animal research: reporting in vivo experiments: The ARRIVE guidelines », 2010

is the most cited work (3465)

Top 10 Works - BJP Journal



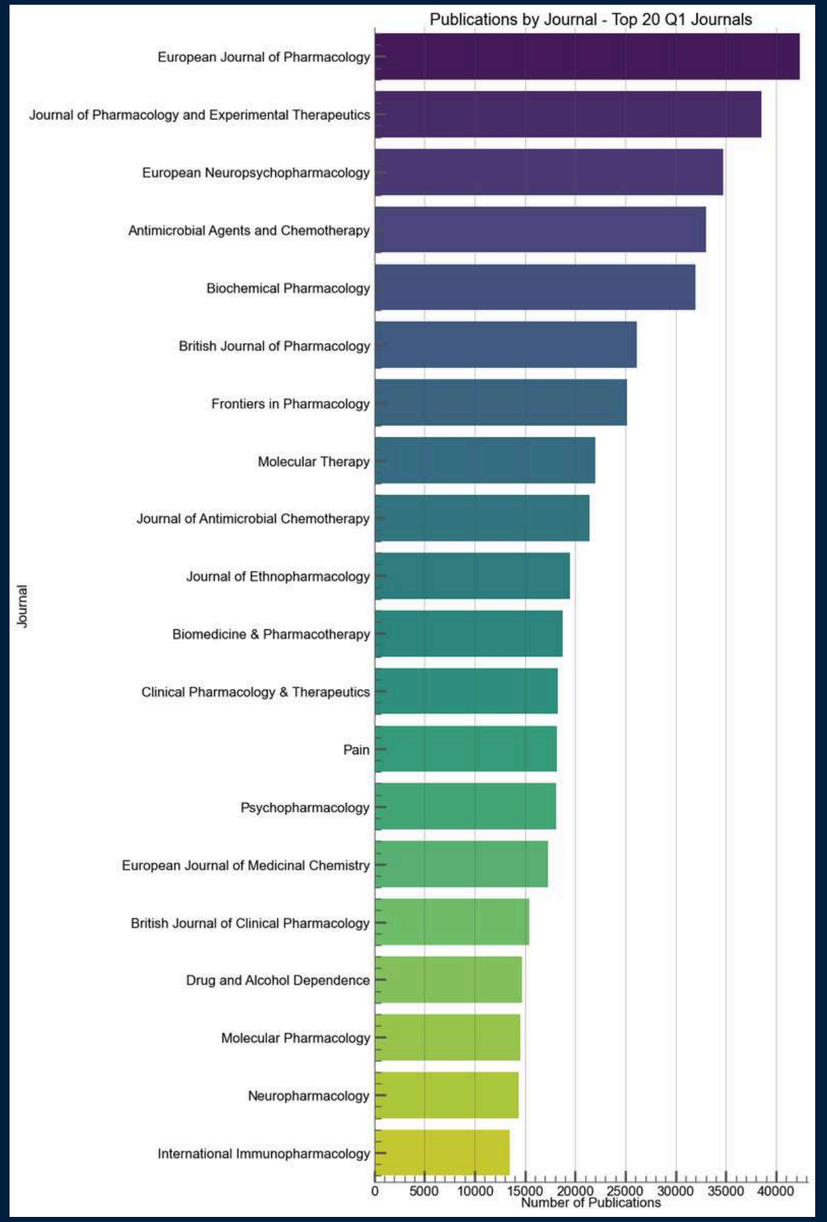
HSDSLab
Human & Social Data Science

General Metrics

Publications - Journal Ranking

BJP is ranked **6th**, with these

26 115 publications !



HSDSLab
Human & Social Data Science

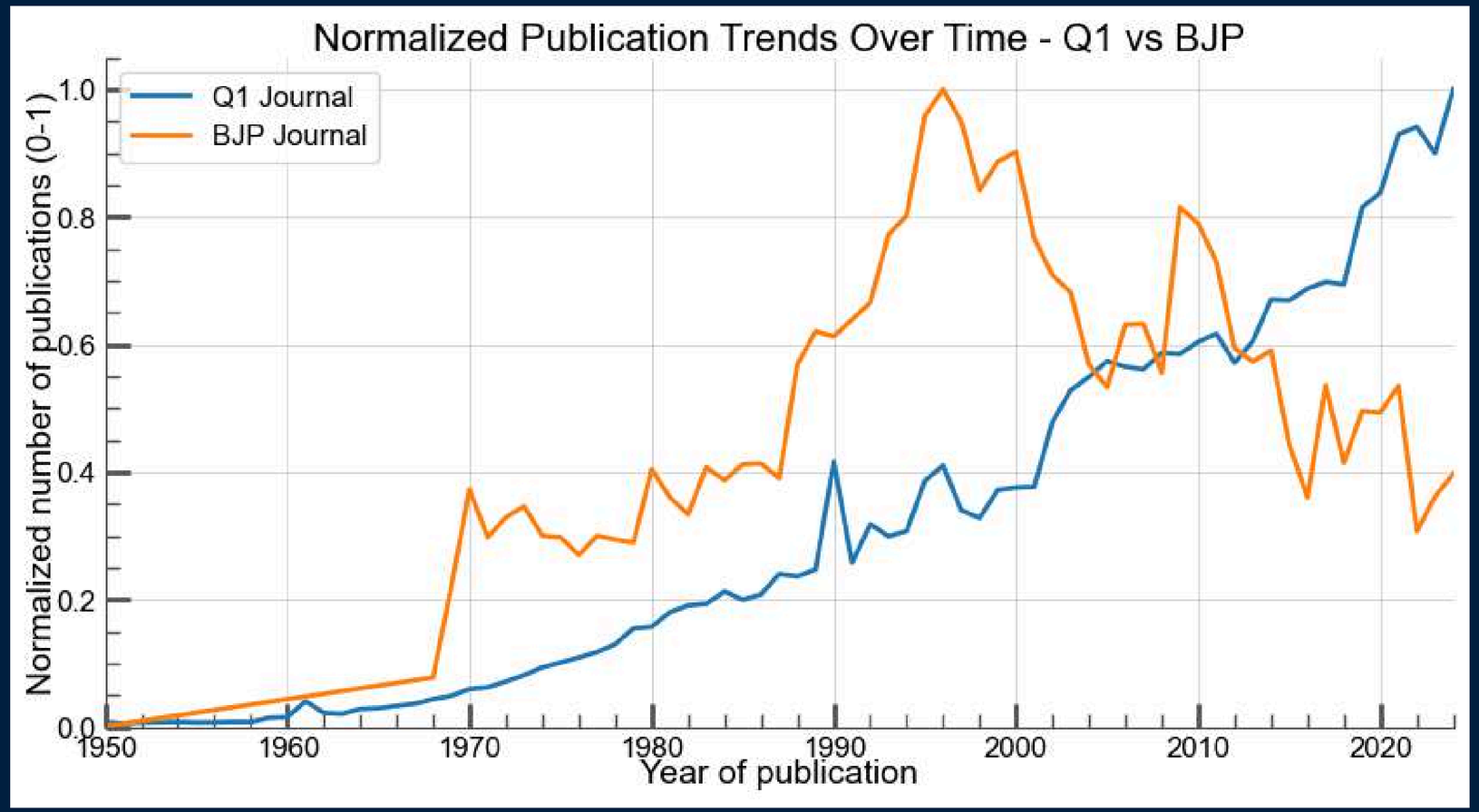
General Metrics

Publications - Works

- *Normalized* : Values are scaled relative to the yearly maximum, ranging from 0 to 1.

- Peak in 1996 for the BJP (864 Publications)

- Peak in 2024 for Q1 Journals (29378 Publications)



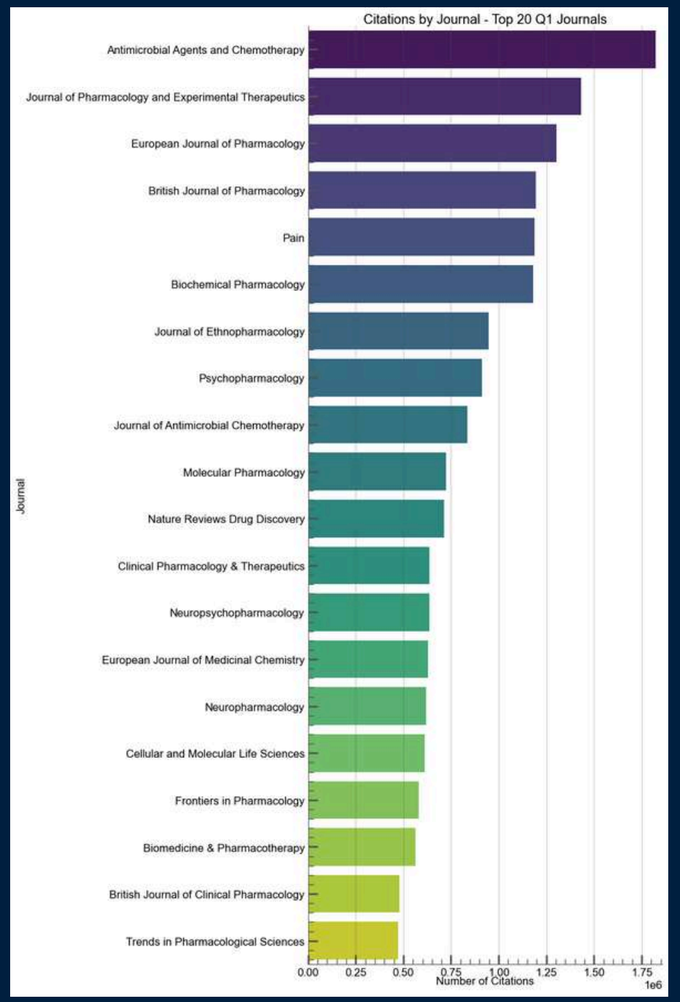
General Metrics

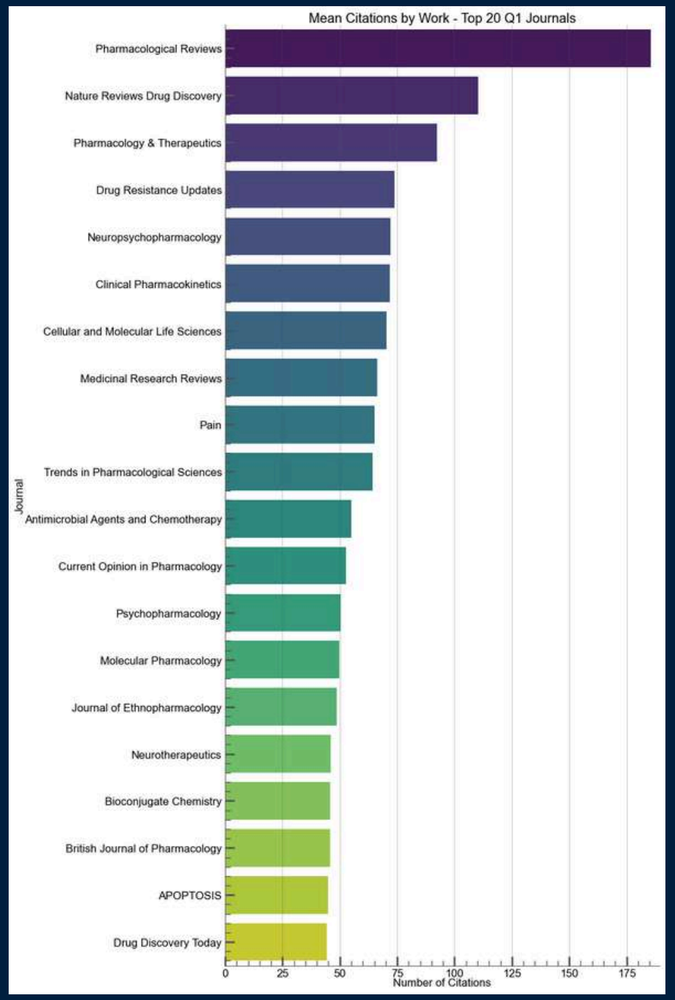
Citations - Based on the year of the work cited
- Journal Ranking

- Cumulative : BJP is ranked **4th**,

  with 1 193 827 total citations !

- Mean : BJP is ranked **18th**,
  with around 45.17 citations
  per work.

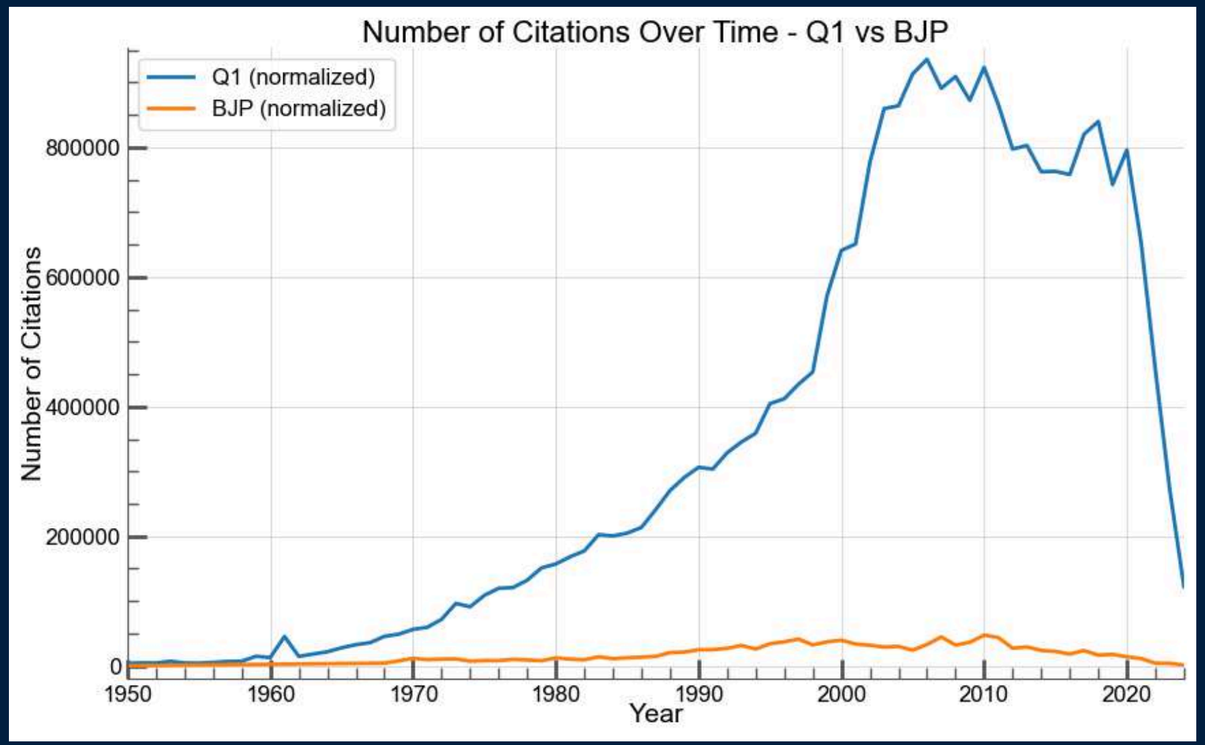


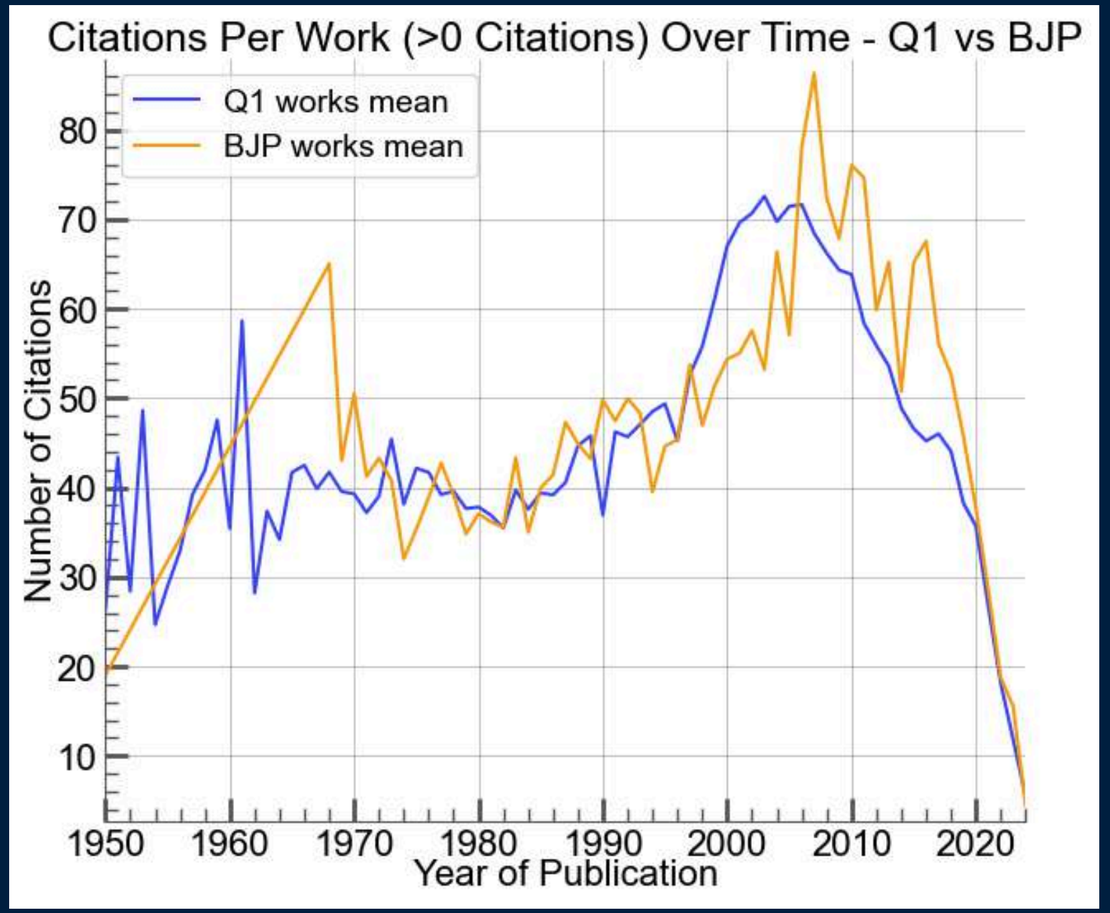


HSDSLab
Human & Social Data Science

General Metrics

Citations - Based on the year of the work cited





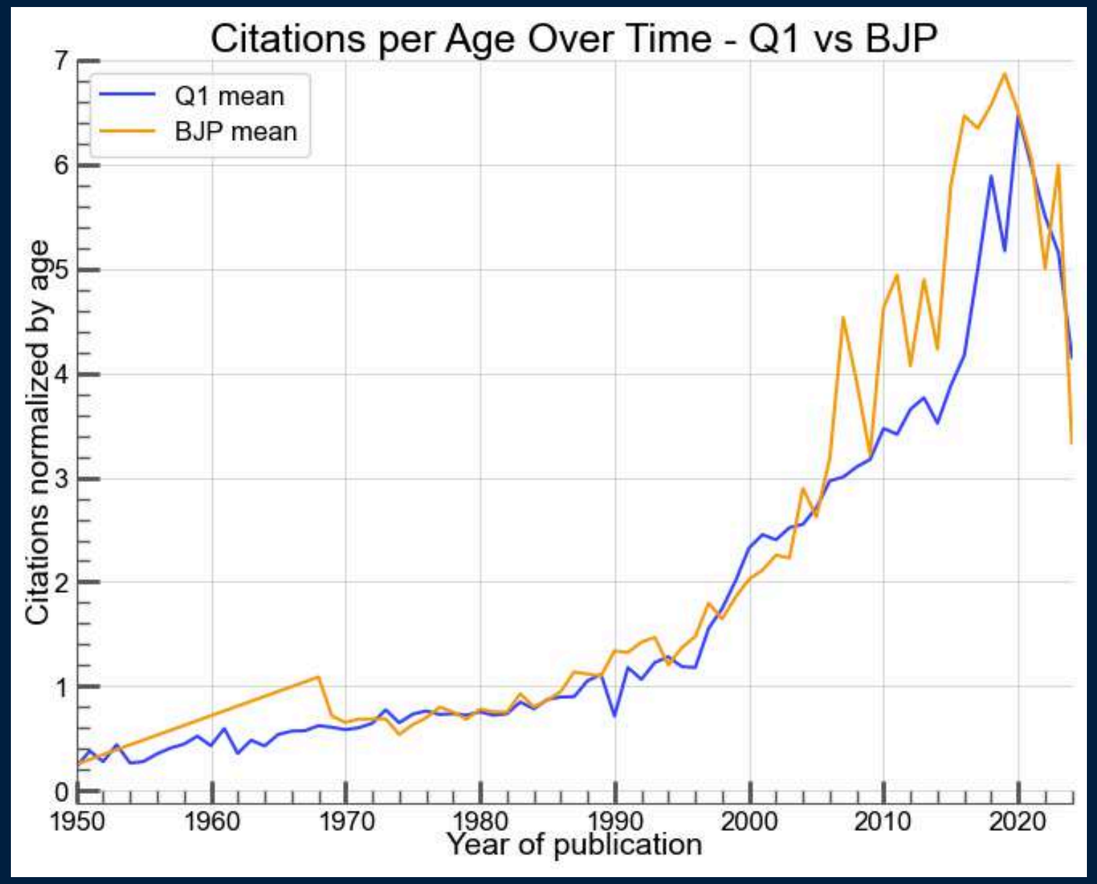
Peak in 2010 for BJP : 47374 Citations

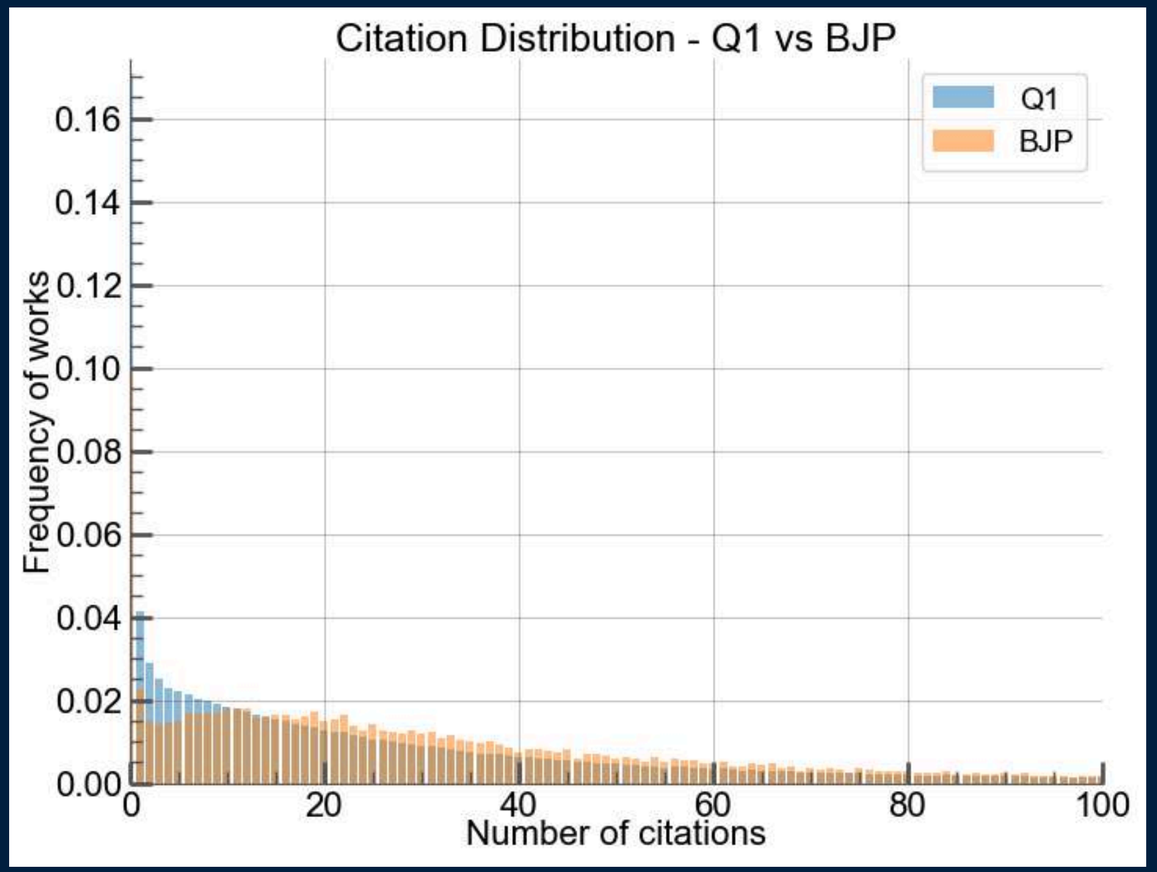
Peak in 2006 for Q1 Journals : 935948 Citations

HSDSLab
Human & Social Data Science

General Metrics

Citations - Based on the year of the work cited

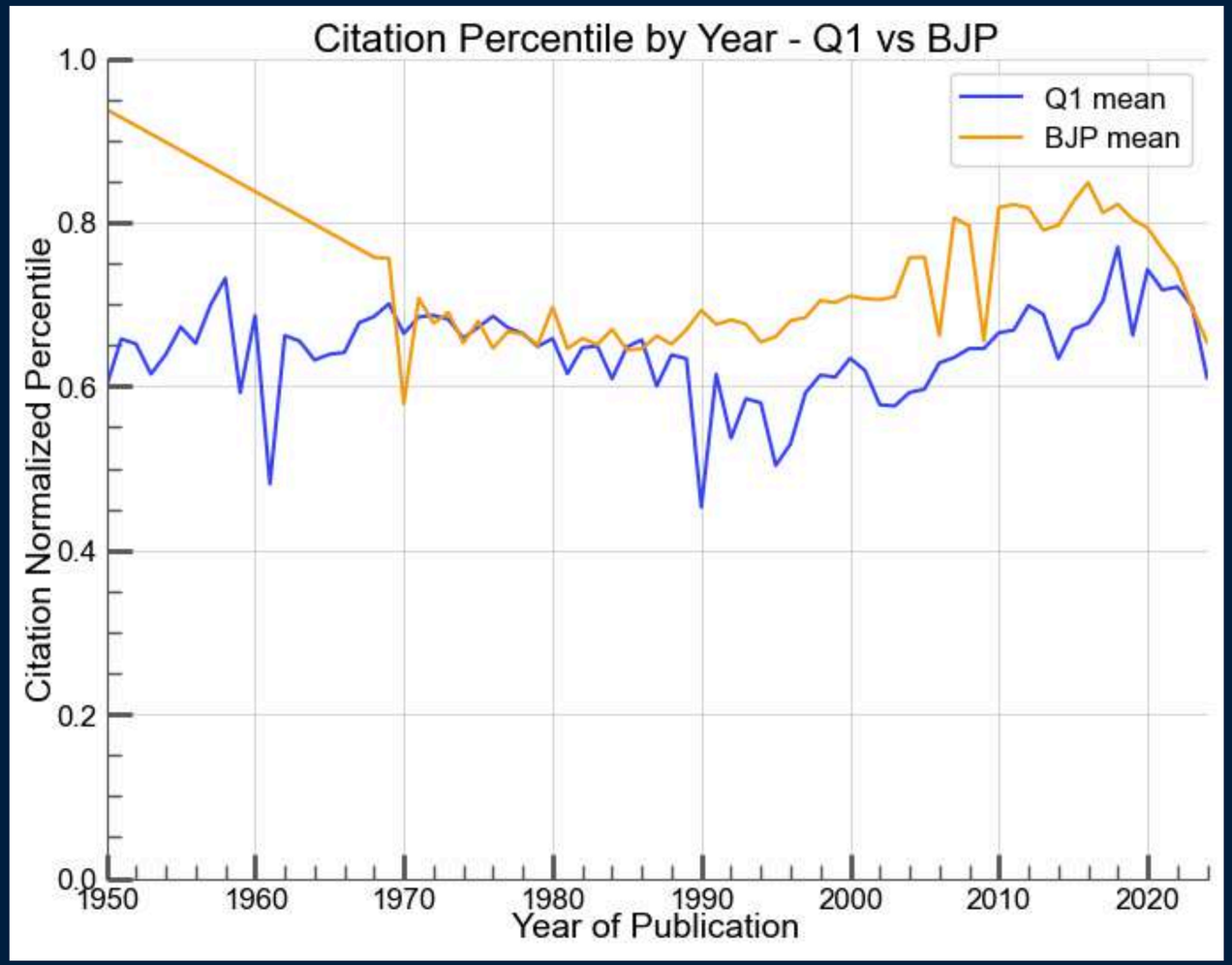




General Metrics

Citations - The Citation Percentile metric

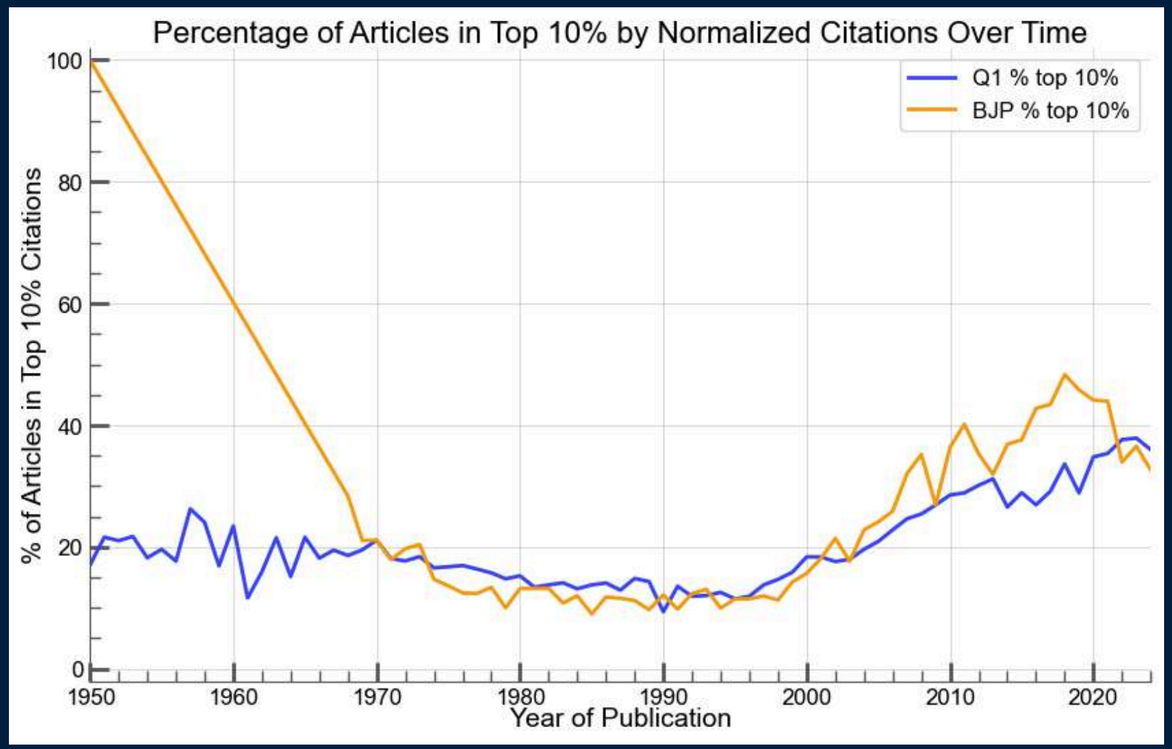
- Metric from OpenAlex
- Calculated among ALL pharmacology articles from the same year.
- Shows how much a paper is cited compared to similar papers.



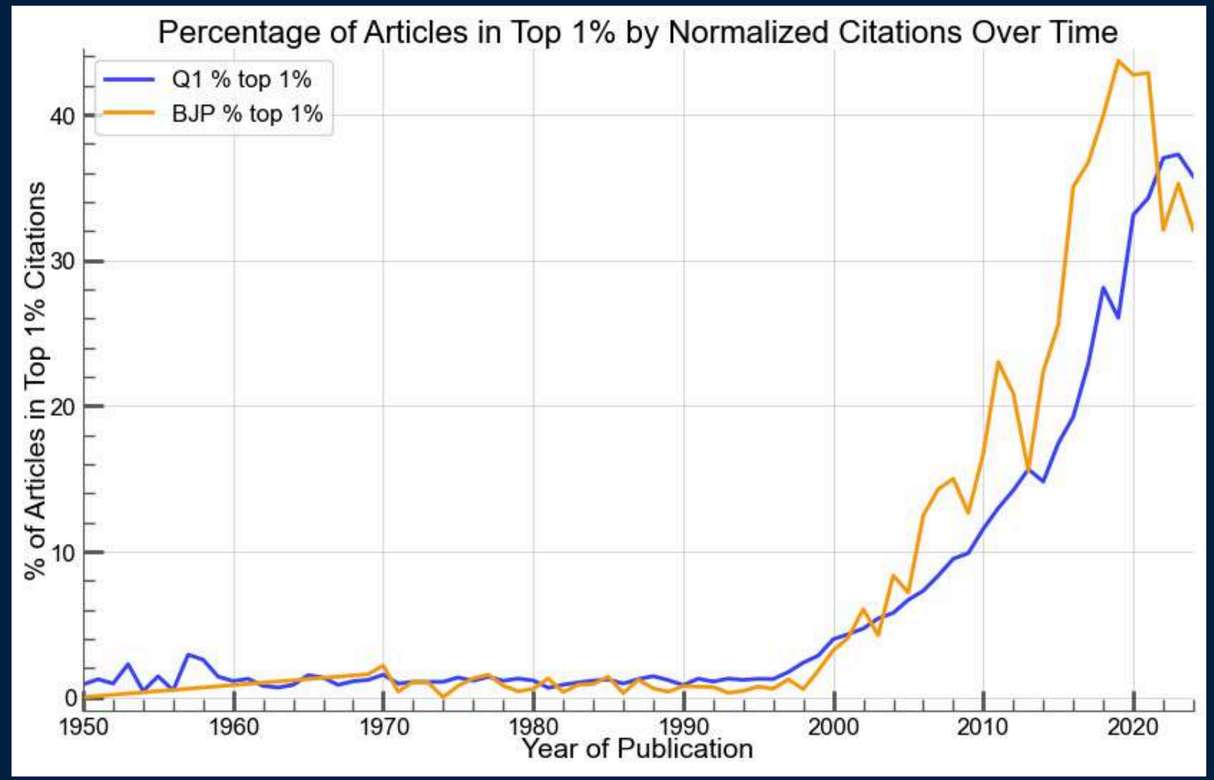
HSDSLab
Human & Social Data Science

General Metrics

Citations - The Citation Percentile metric



Top 10 % : Citation Percentile >= 0.9



Top 1 % : Citation Percentile >= 0.99

HSDSLab
Human & Social Data Science

General Metrics

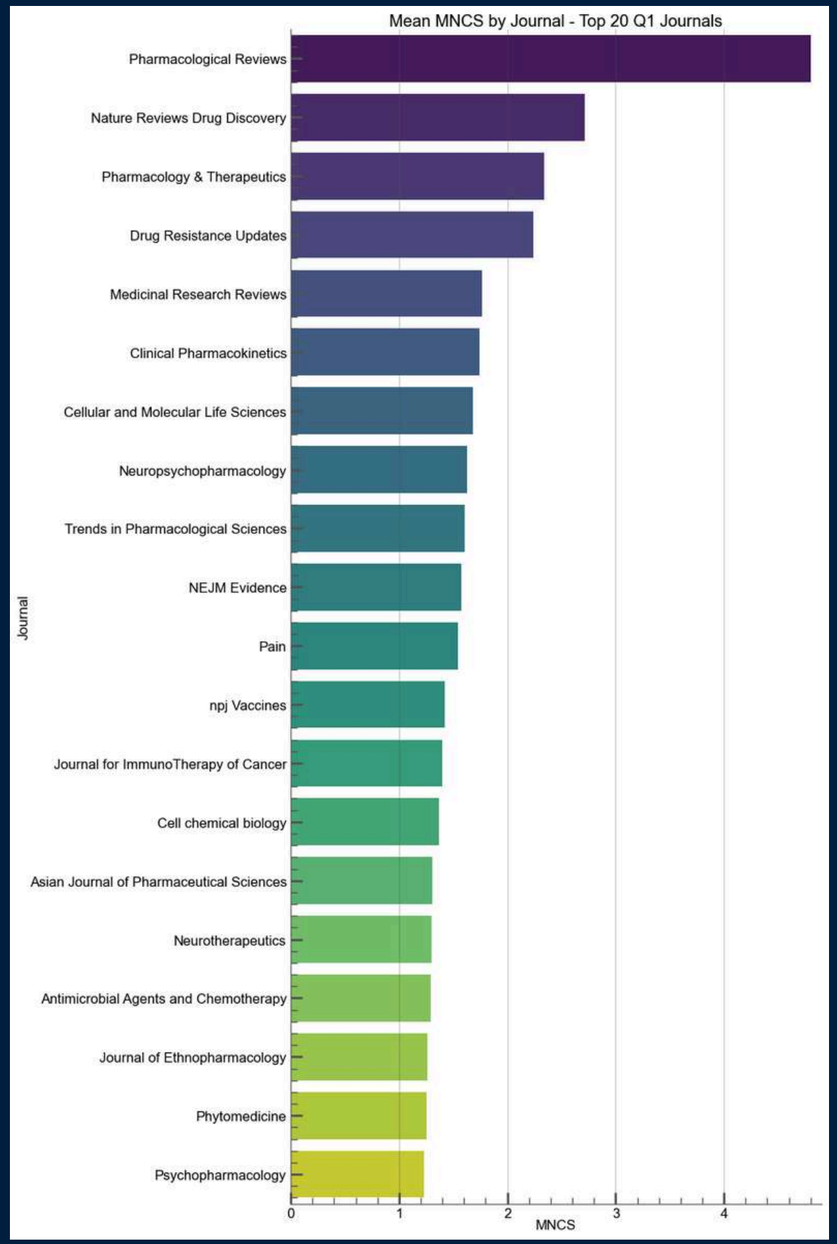
The MNCS Metric - Journal Ranking

- MNCS = Mean Normalized Citation Score

$$\frac{\text{Citations of the work}}{\text{Mean Citations for same year works}}$$

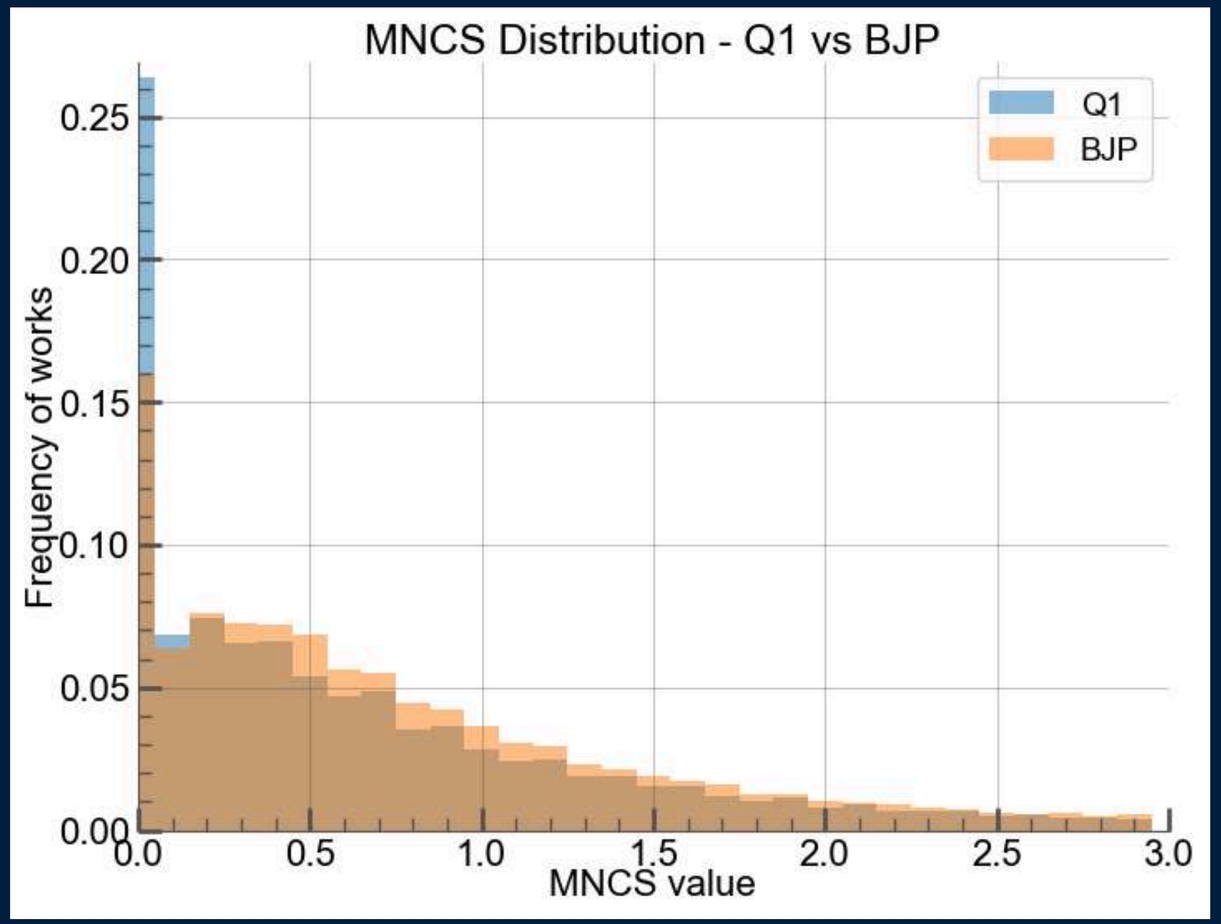
BJP is ranked **25th**, with around

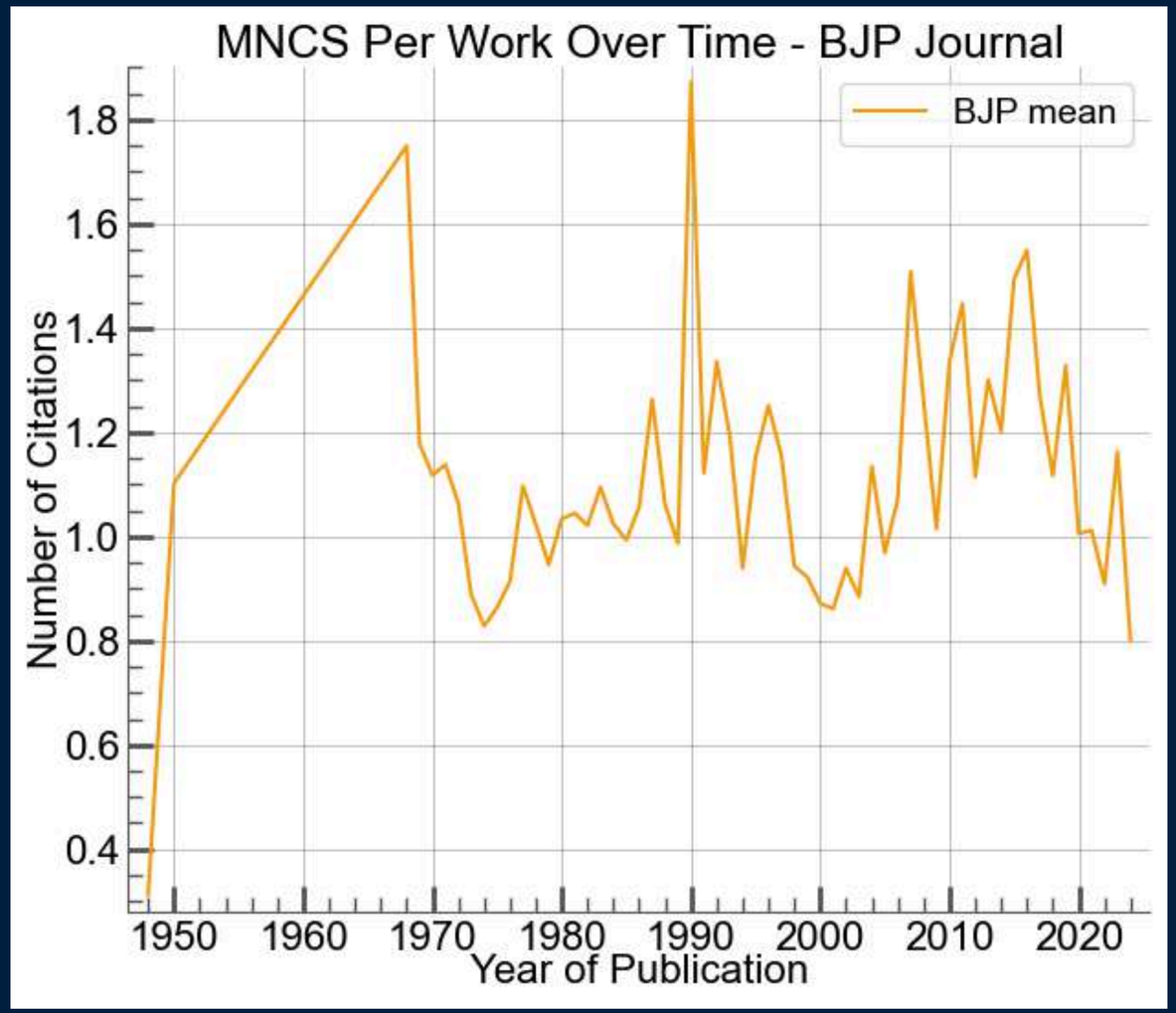
1.12 MNCS per work.



General Metrics

Citations - The MNCS metric



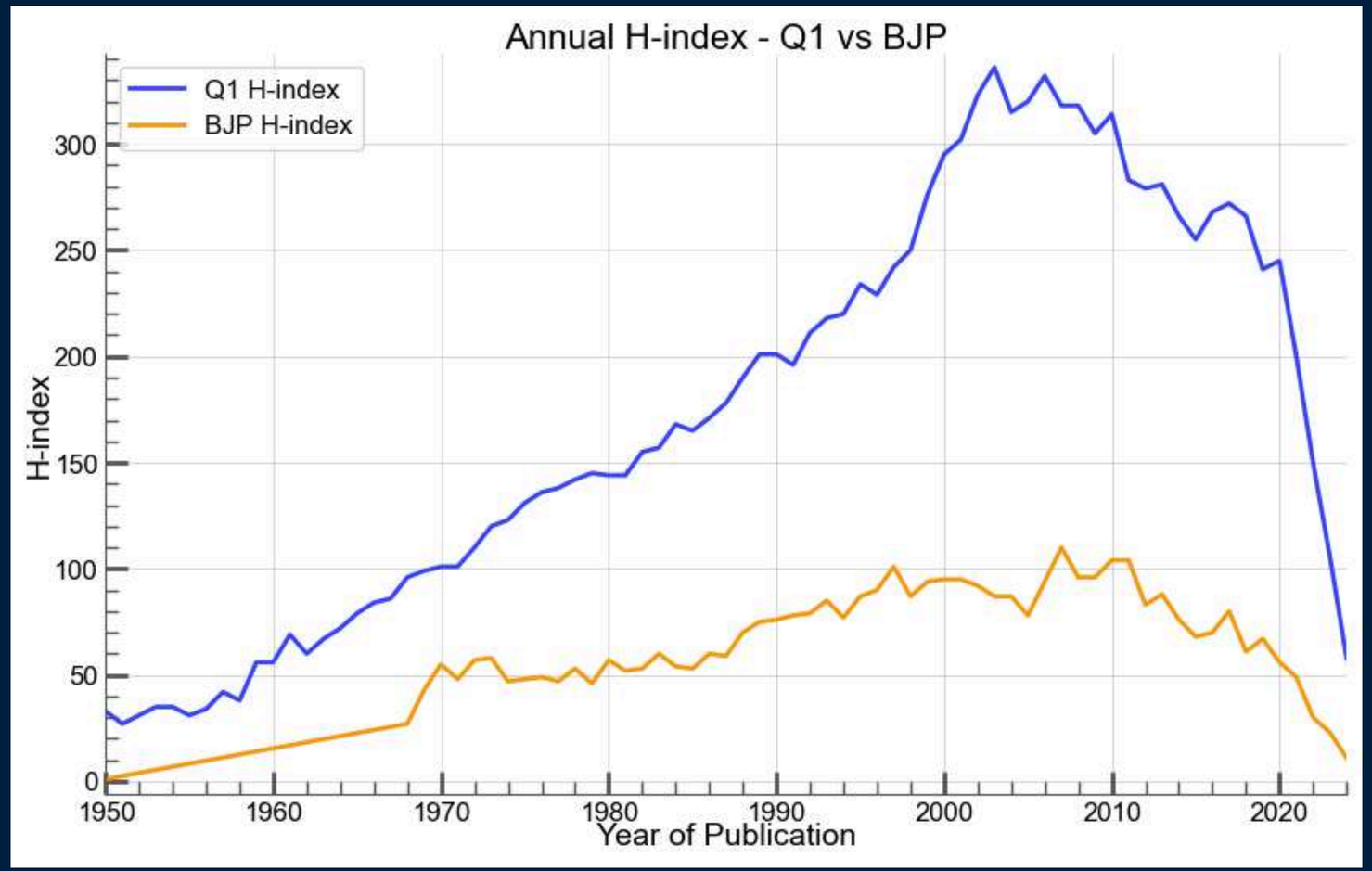


HSDSLab
Human & Social Data Science

General Metrics

H-Index

- Measures both productivity and impact or journal.
- H = Number of papers that have at least H citations each.
- Similar trend, but Q1 (336 works with 336+ citations in 2003) reaches a much higher peak than BJP (110 works with 110+ citations in 2007)



HSDSLab
Human & Social Data Science

General Metrics

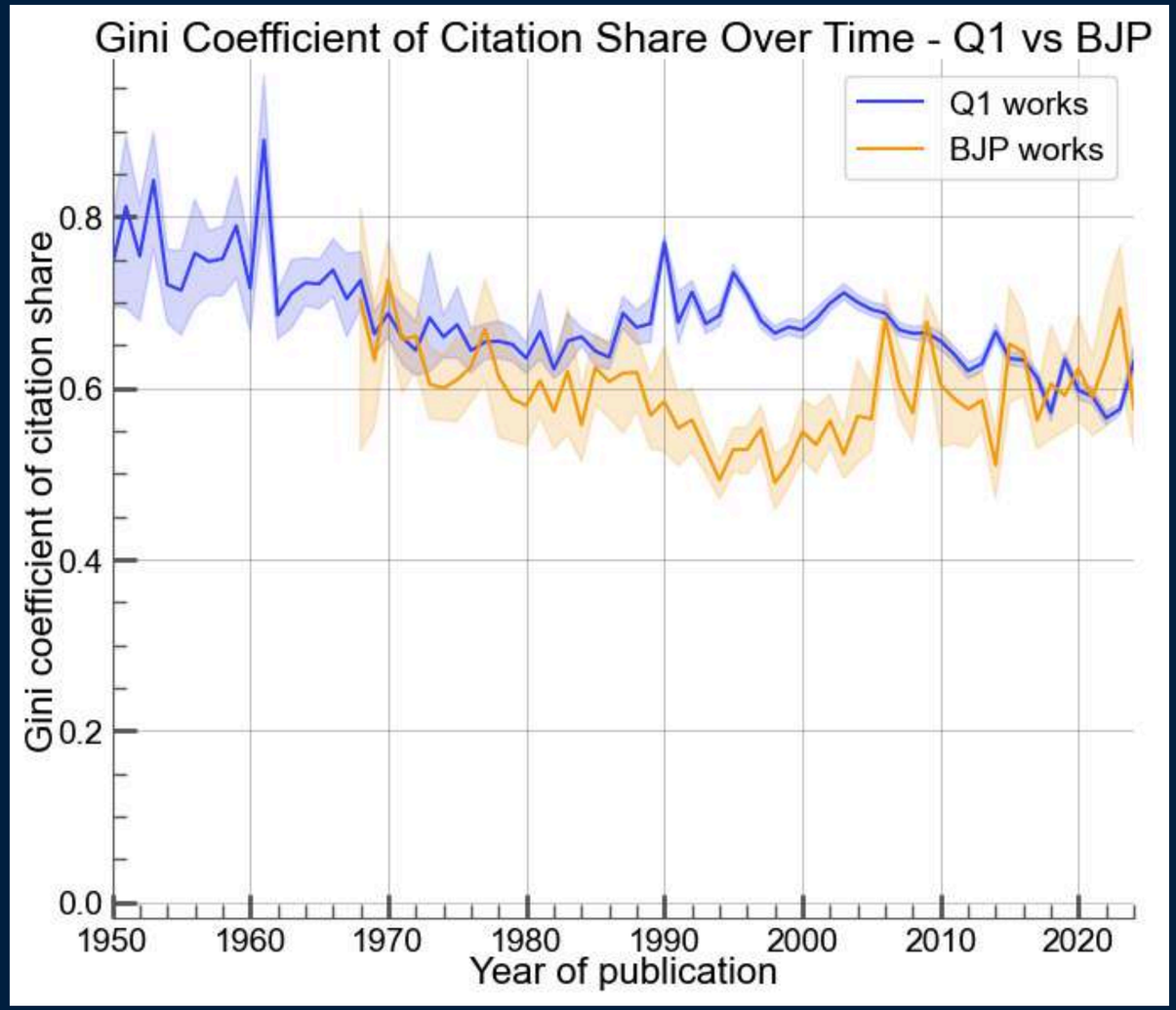
Gini Coefficient - Citations Share

- Measures inequality in citation distribution

- 0 = Perfect equality (all papers cited equally)
- 1 = Maximum inequality
  (few papers get most citations)

- Gini coefficient was lower (by ~0.1–0.2) between 1985 and 2010 for the BJP

→ This indicates a more equal distribution during that period



HSDSLab
Human & Social Data Science

General Metrics

Citations since 2012 - Journal Ranking

Unlike the previous citation metric, this one reflects the citations received by the journal year by year (from 2012 onward), regardless of the publication year of the cited articles.

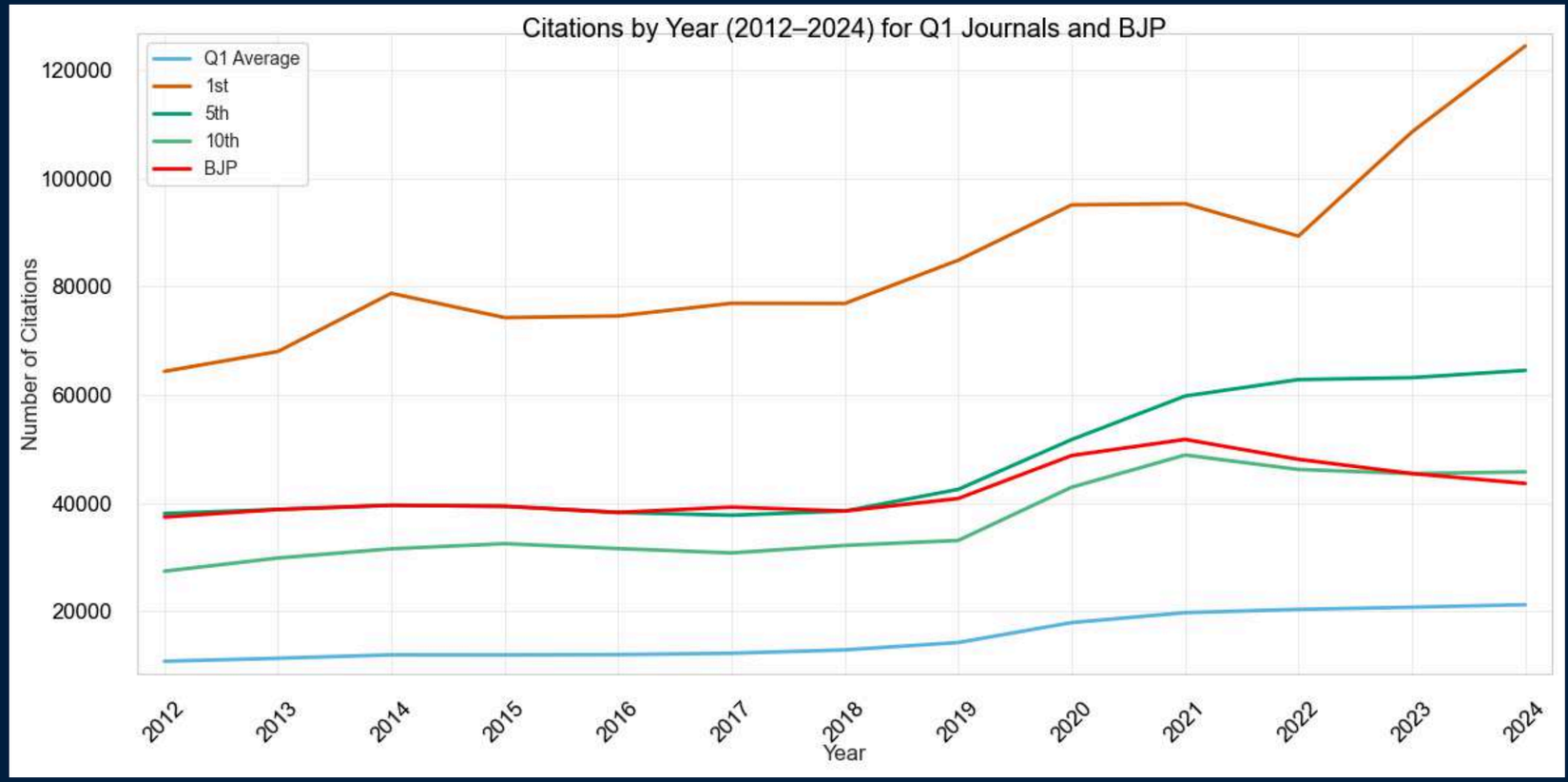
BJP is ranked **6th**, with

220 058 total citations since 2012 !

General Metrics

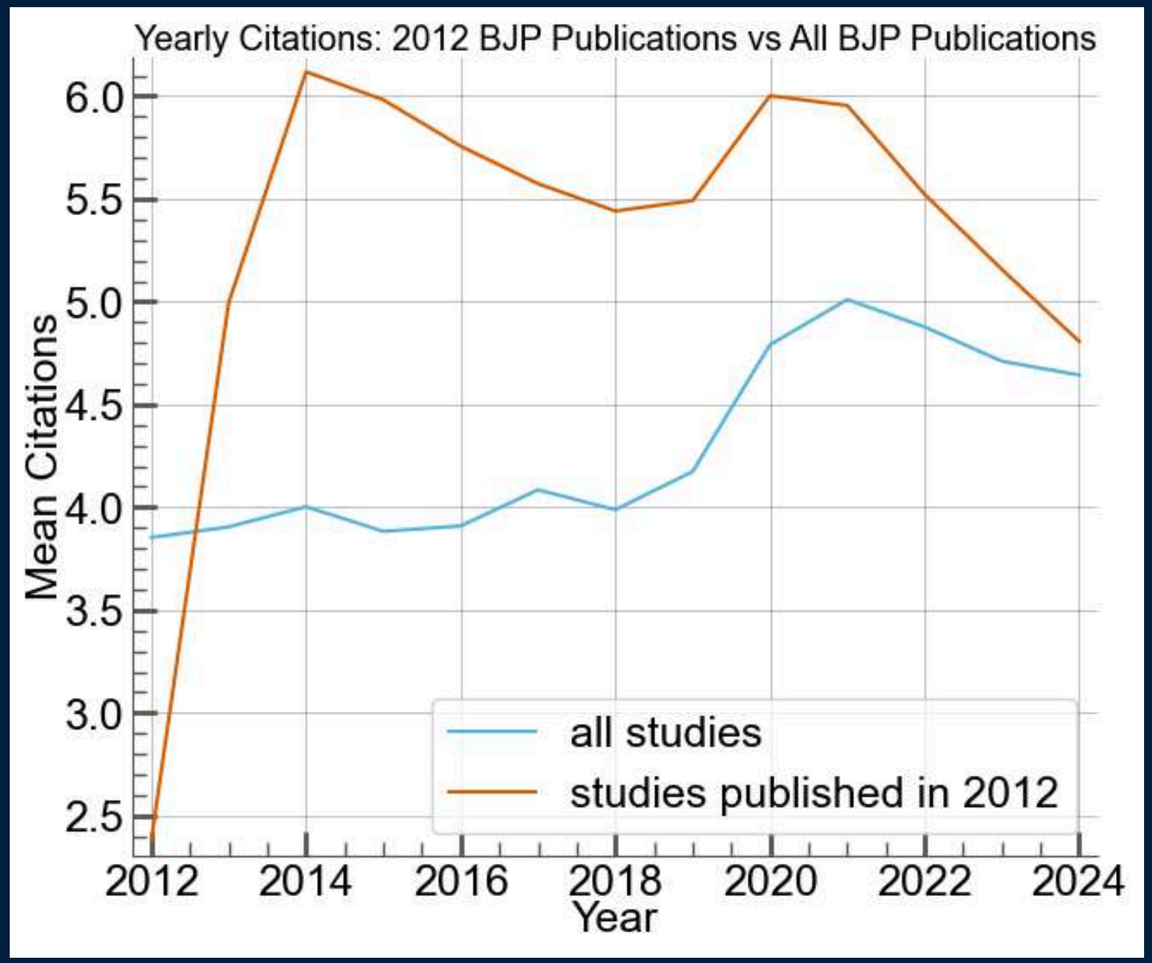
Citations since 2012 - Evolution

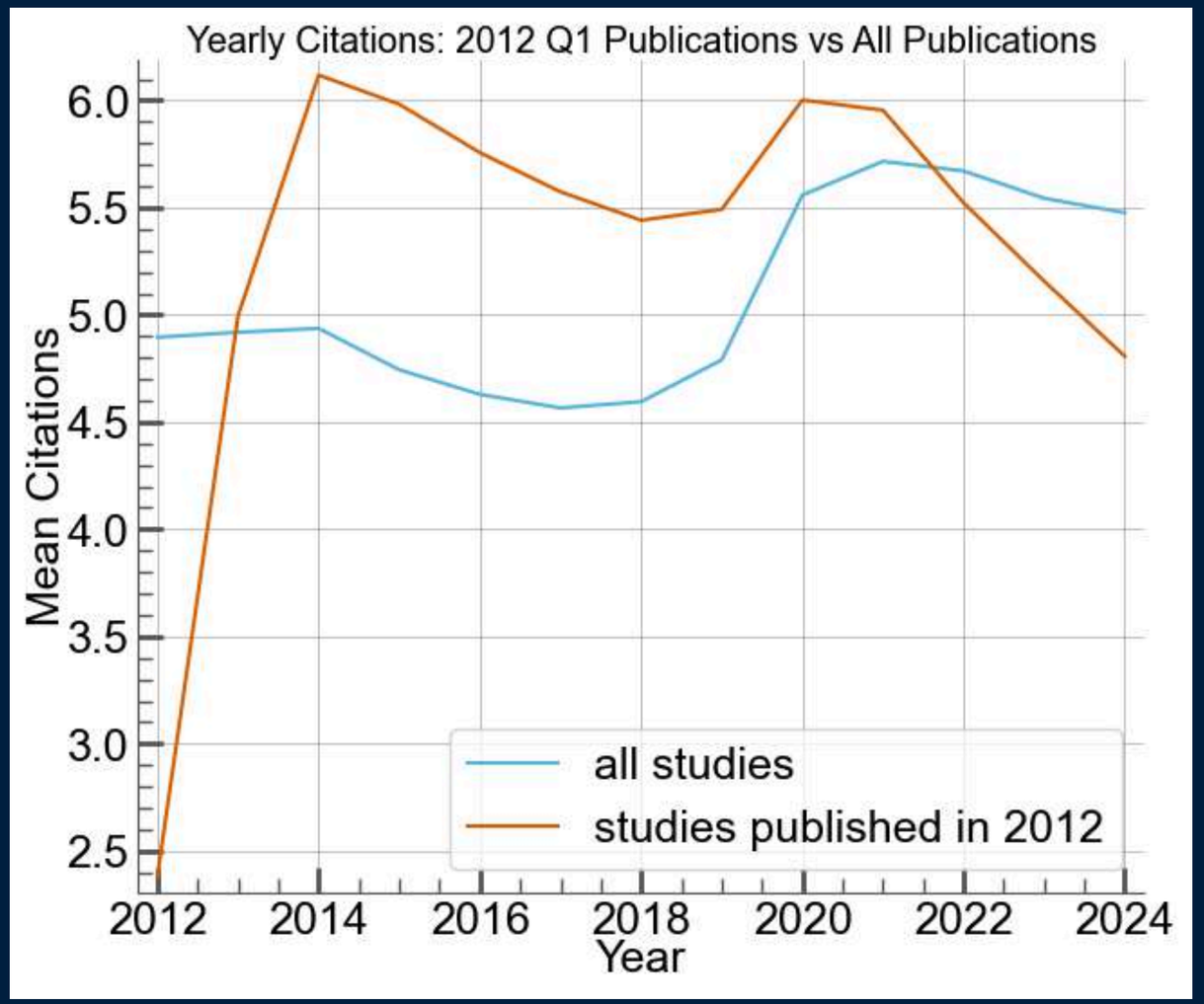
- *Top x* : Number of citations of the x ranked journal this year
- Confirms the ranking, with BJP almost always in the top 5
- Fewer citations since 2021, outside of top 10 in 2024



General Metrics

Citations since 2012 - Age Analysis

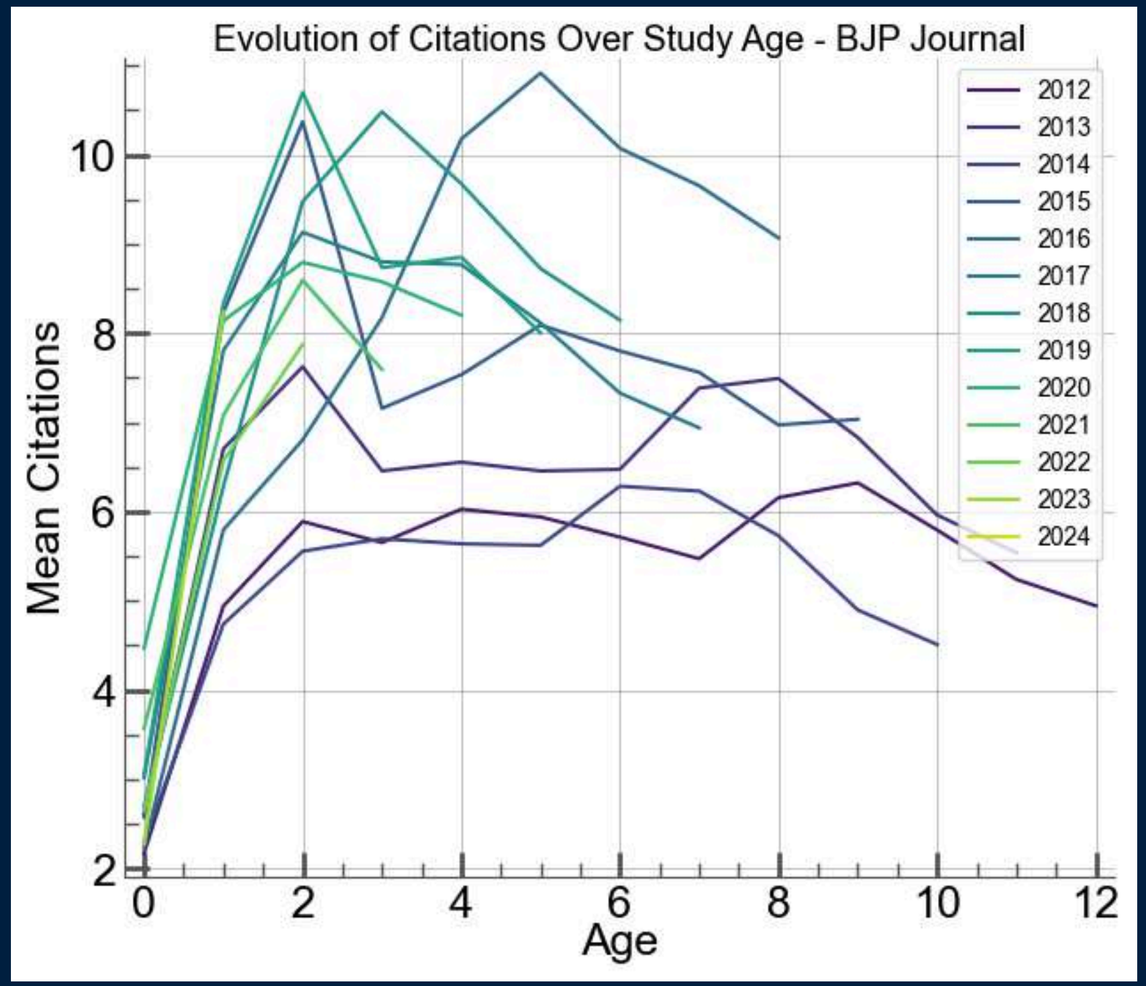


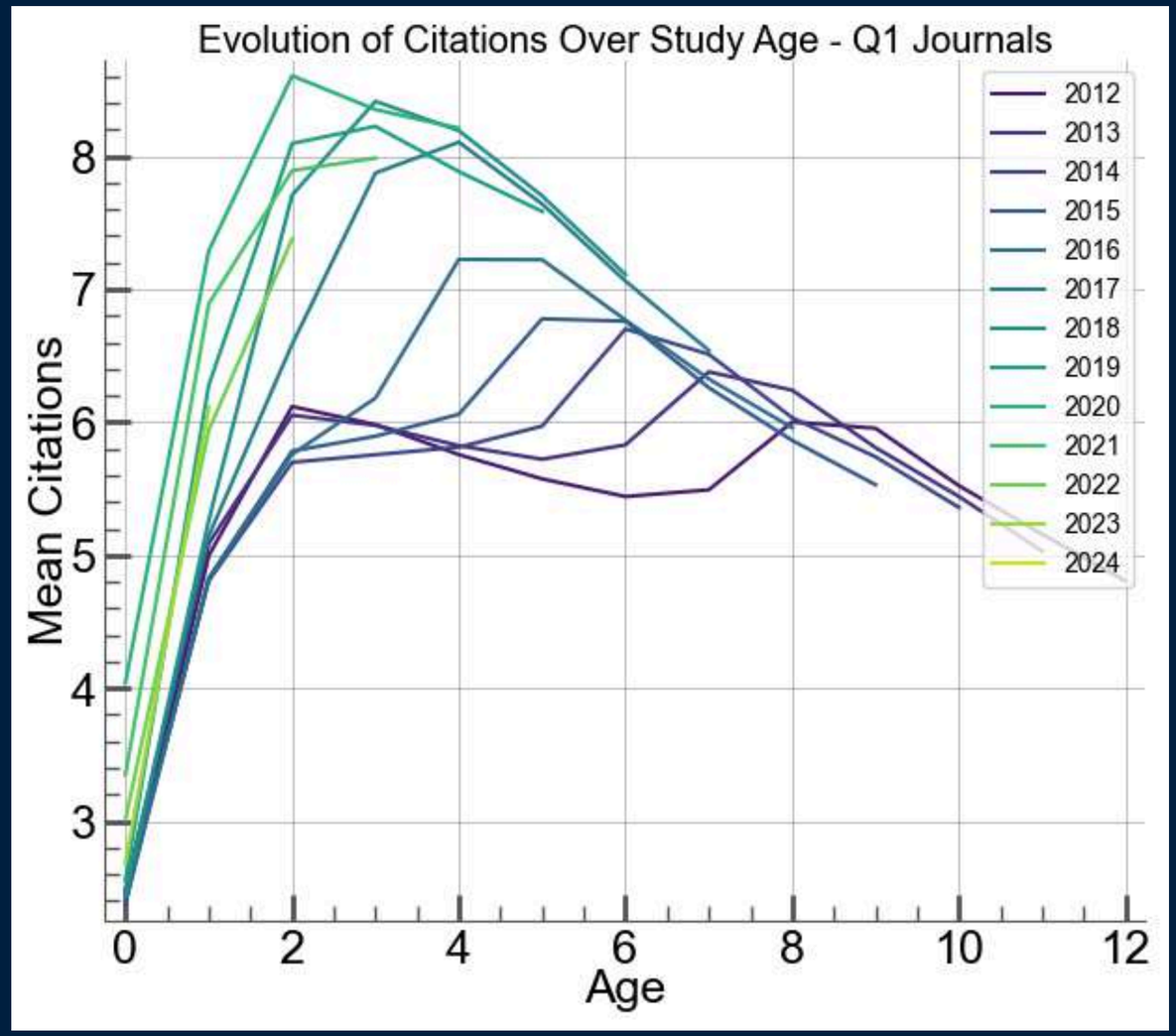


HSDSLab
Human & Social Data Science

General Metrics

Citations since 2012 - Age Analysis



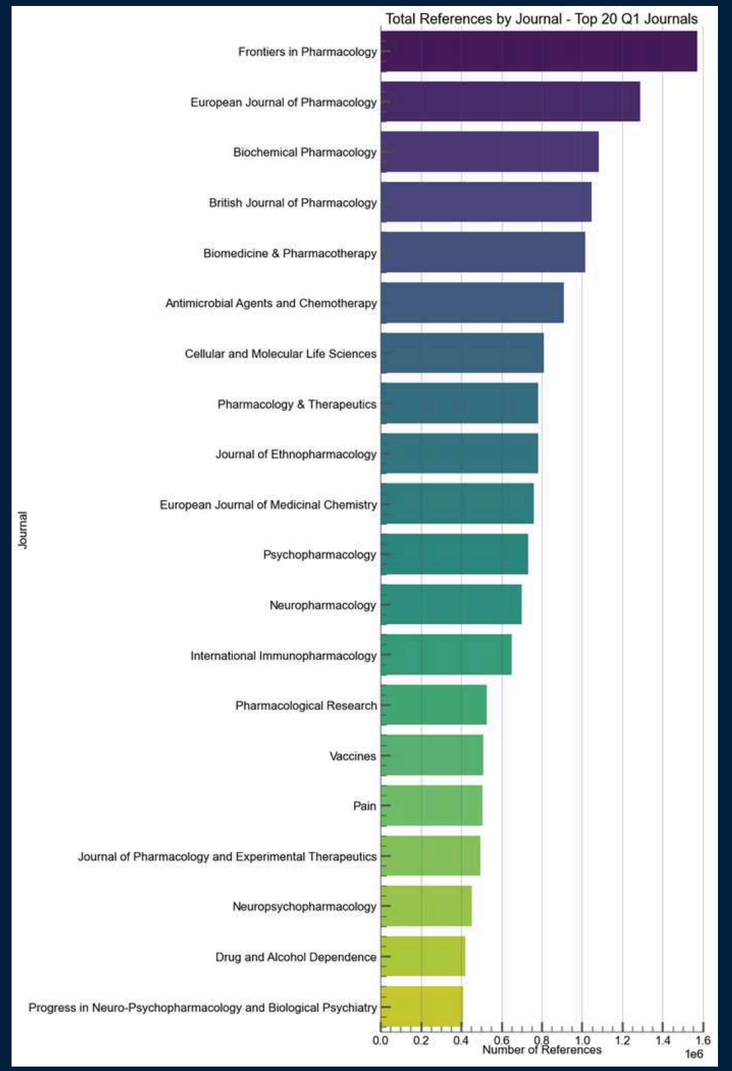


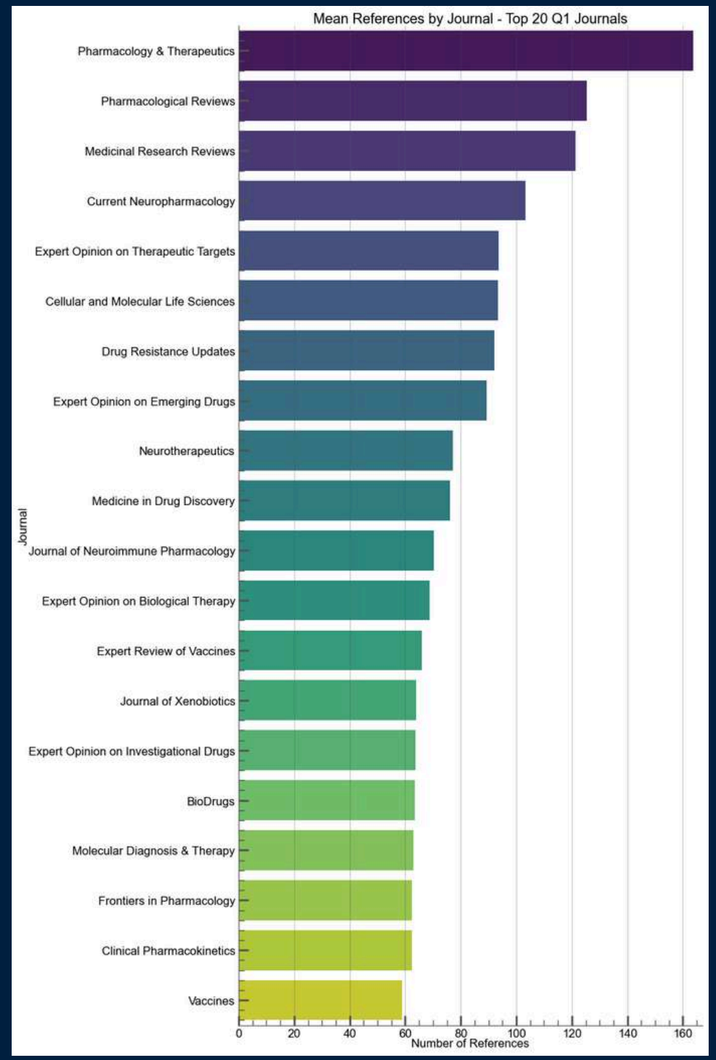
HSDSLab
Human & Social Data Science

General Metrics

Referenced Works - Journal Ranking

- Cumulative : BJP is ranked **4th**, with 1 047 501 total references !
- Mean : BJP is ranked **48th**, with around 40.11 references per work.

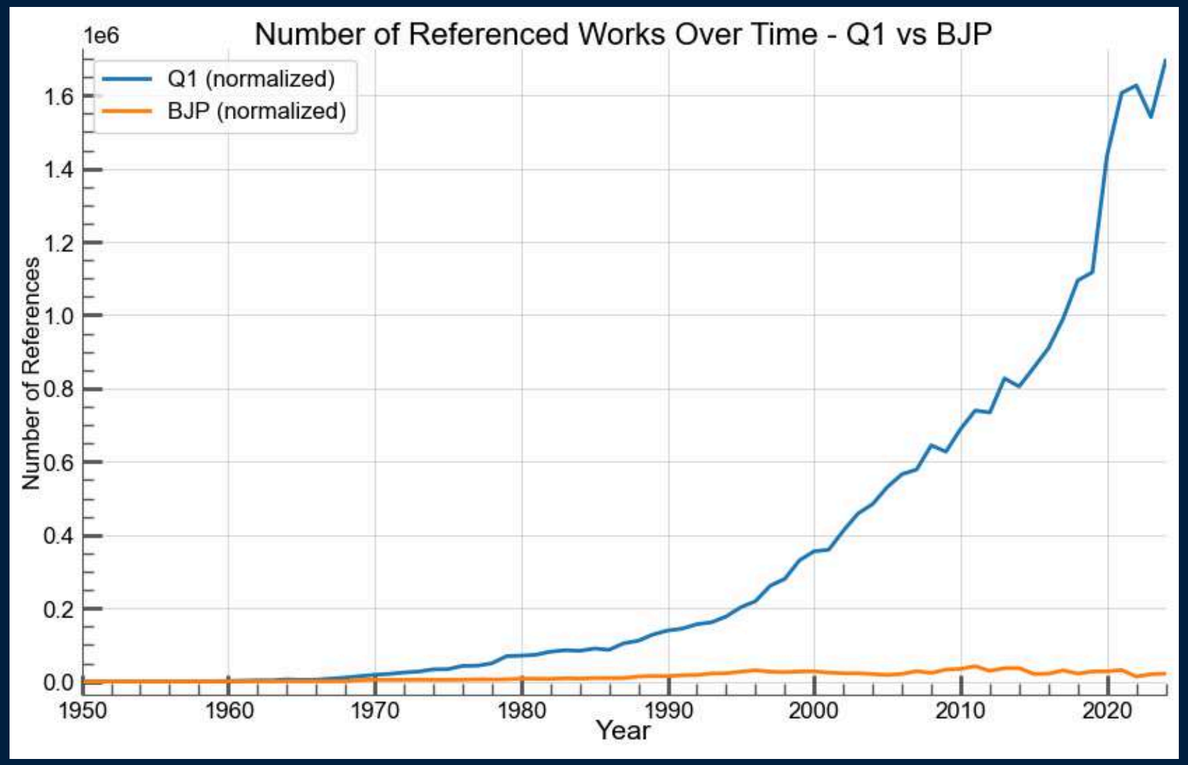


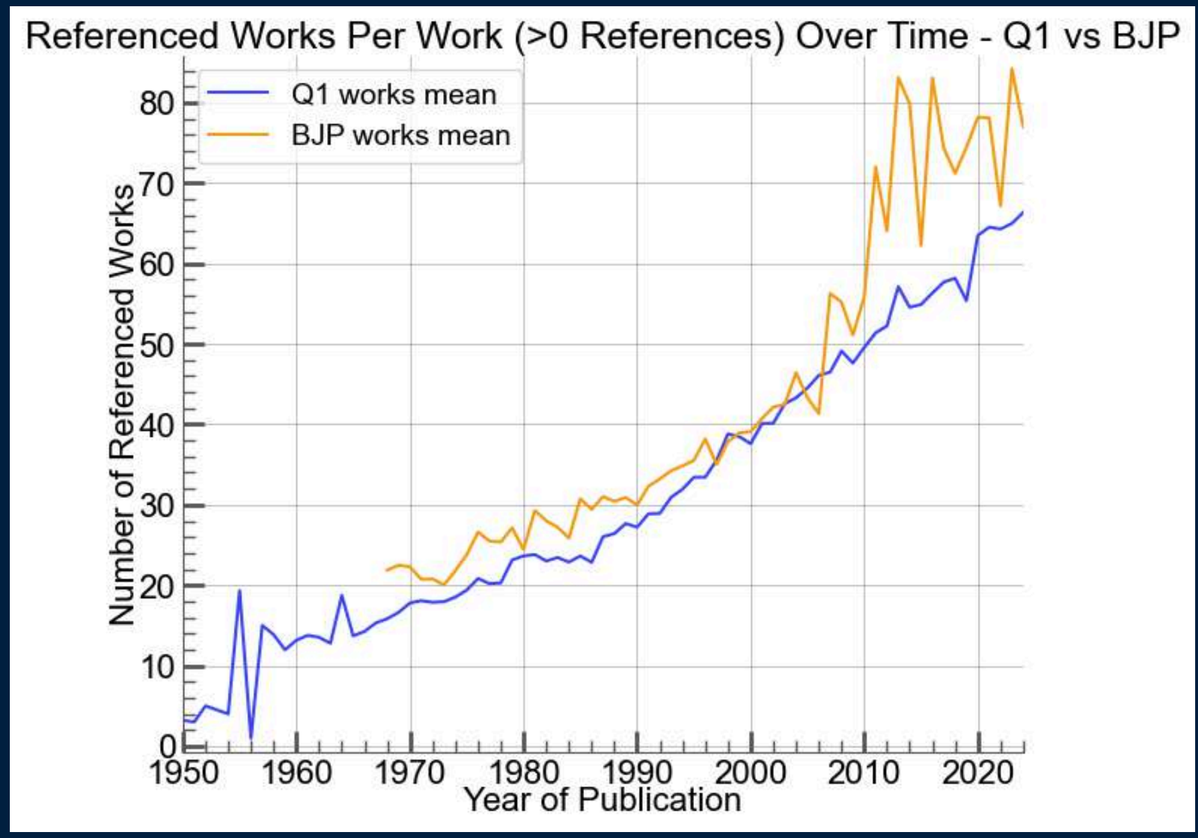


HSDSLab
Human & Social Data Science

General Metrics

Referenced Works





Peak in 2011 for BJP : 41833 References

Peak in 2024 for Q1 Journals : 1694558 References

HSDSLab
Human & Social Data Science

Authors

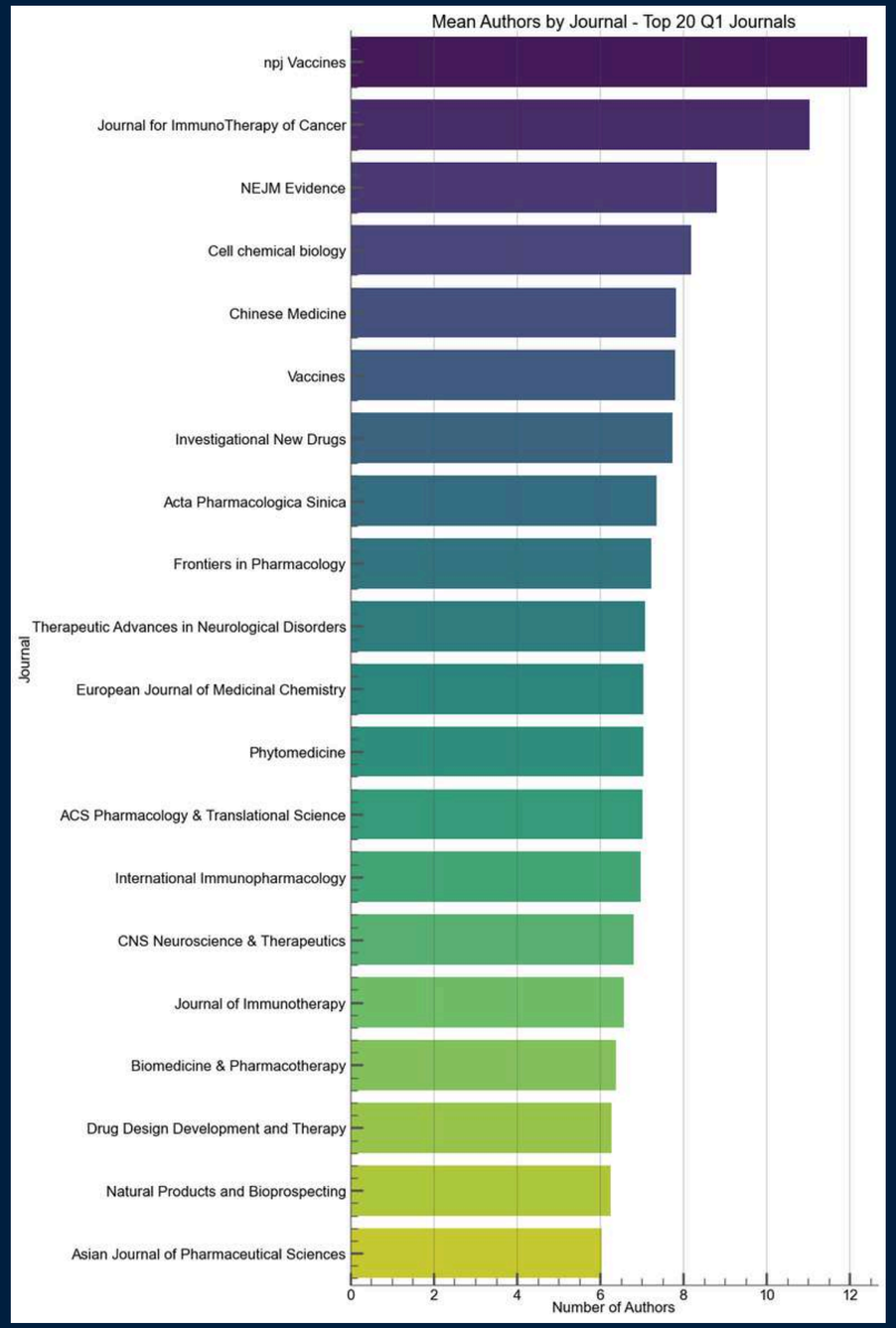
HSDSLab
Human & Social Data Science

Authors

Journal Ranking

BJP is ranked **57th**, with around

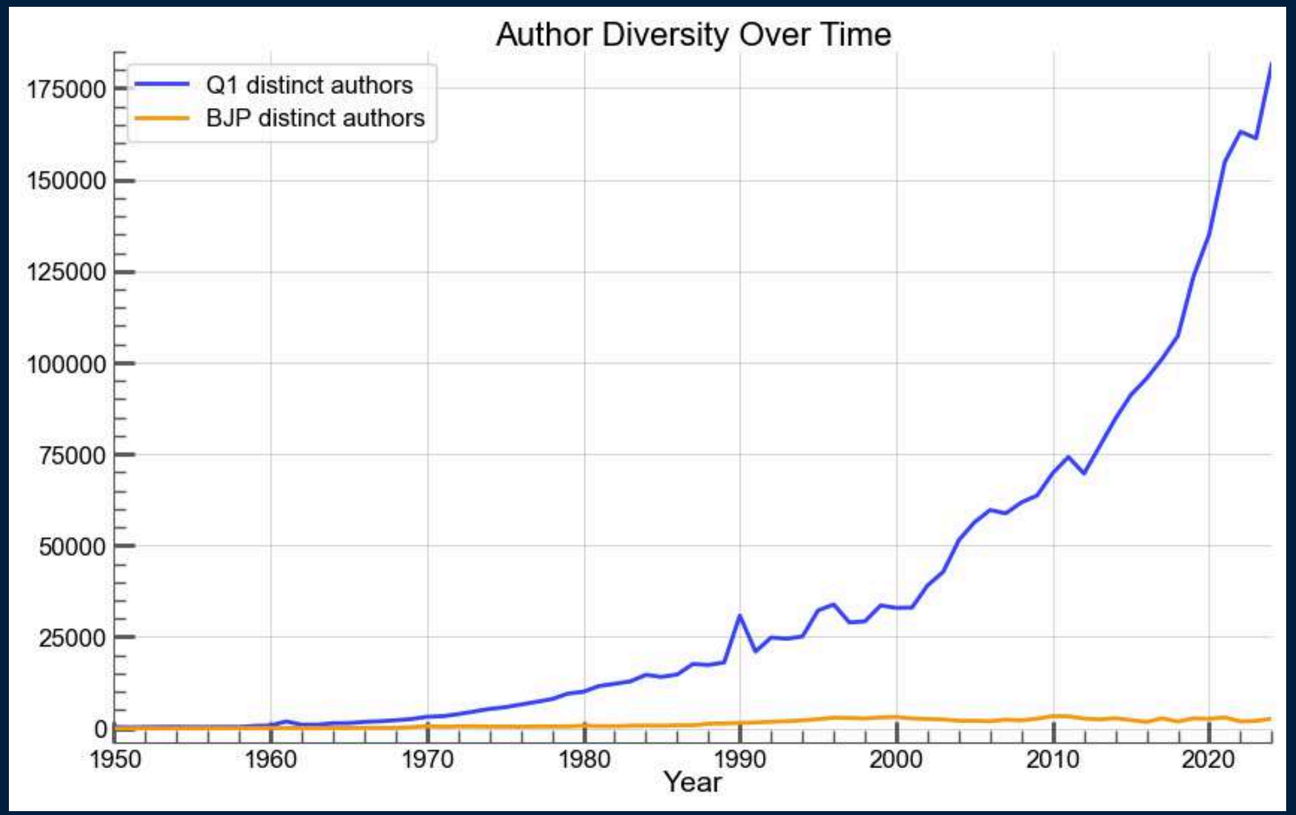
4.29 authors per work.

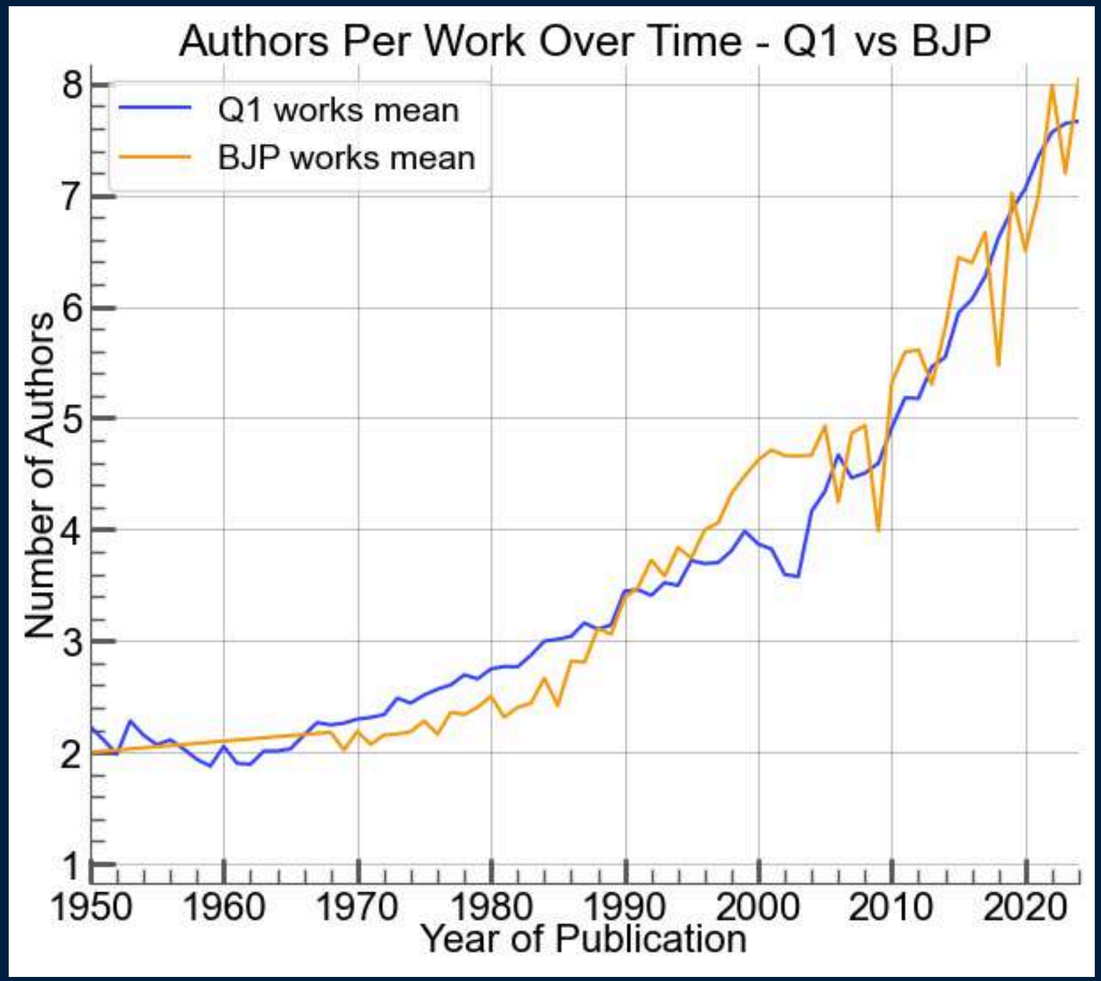


HSDSLab
Human & Social Data Science

Authors

General Metrics





Peak in 2010 for BJP : 3349 Authors

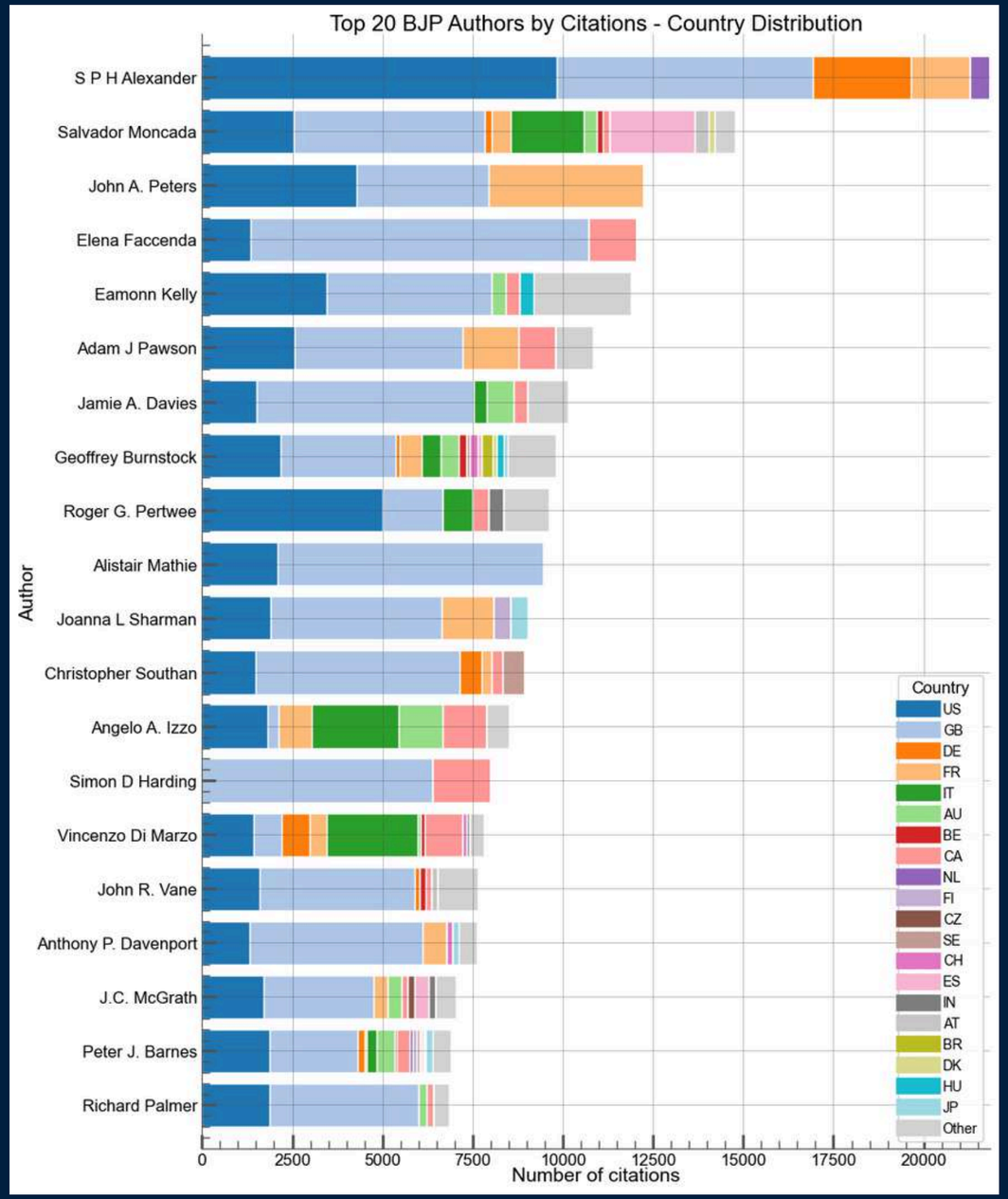
Peak in 2024 for Q1 Journals : 181534 Authors

HSDSLab
Human & Social Data Science

Authors

Top Authors - Top 20 cited authors -BJP Journal

Author countries are based on the countries of their institutions.
For example, S. P. H. Alexander is British

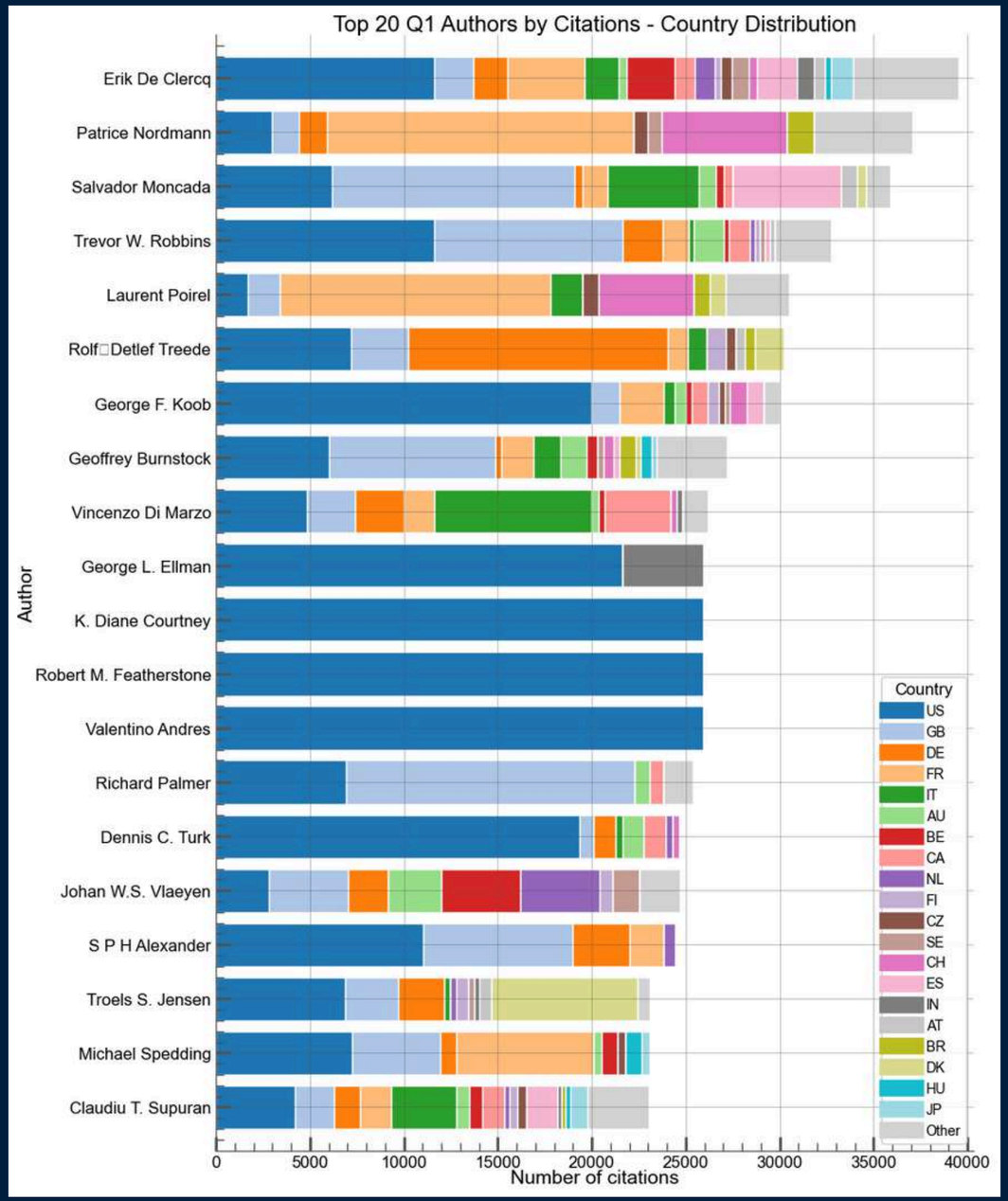


HSDSLab
Human & Social Data Science

Authors

Top Authors - Top 20 cited authors - Q1 Journals

Author countries are based on the countries of their institutions.
For example, Erik De Clercq is Belgian

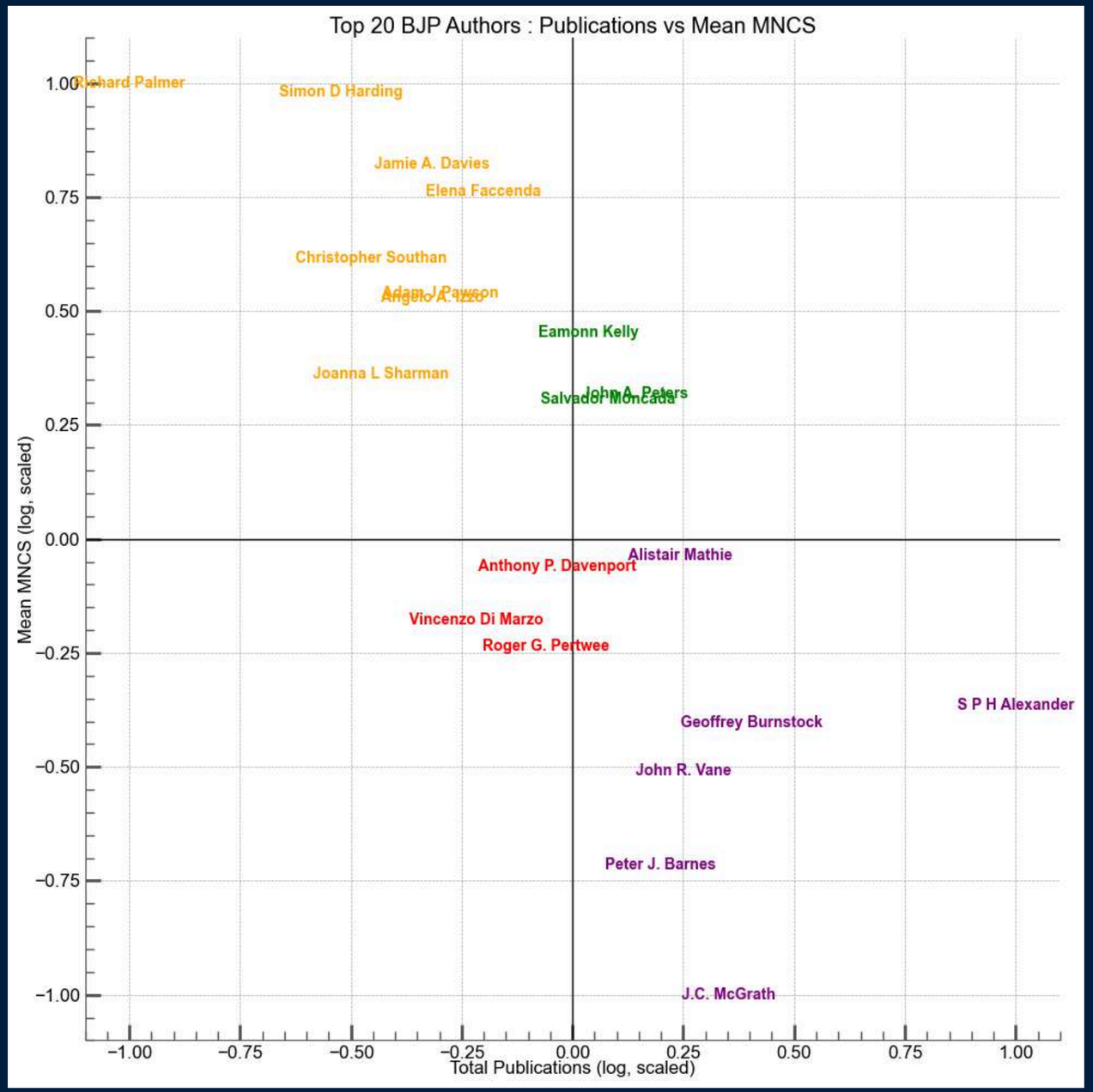


HSDSLab
Human & Social Data Science

Authors

Top Authors - BJP Publications vs Citations

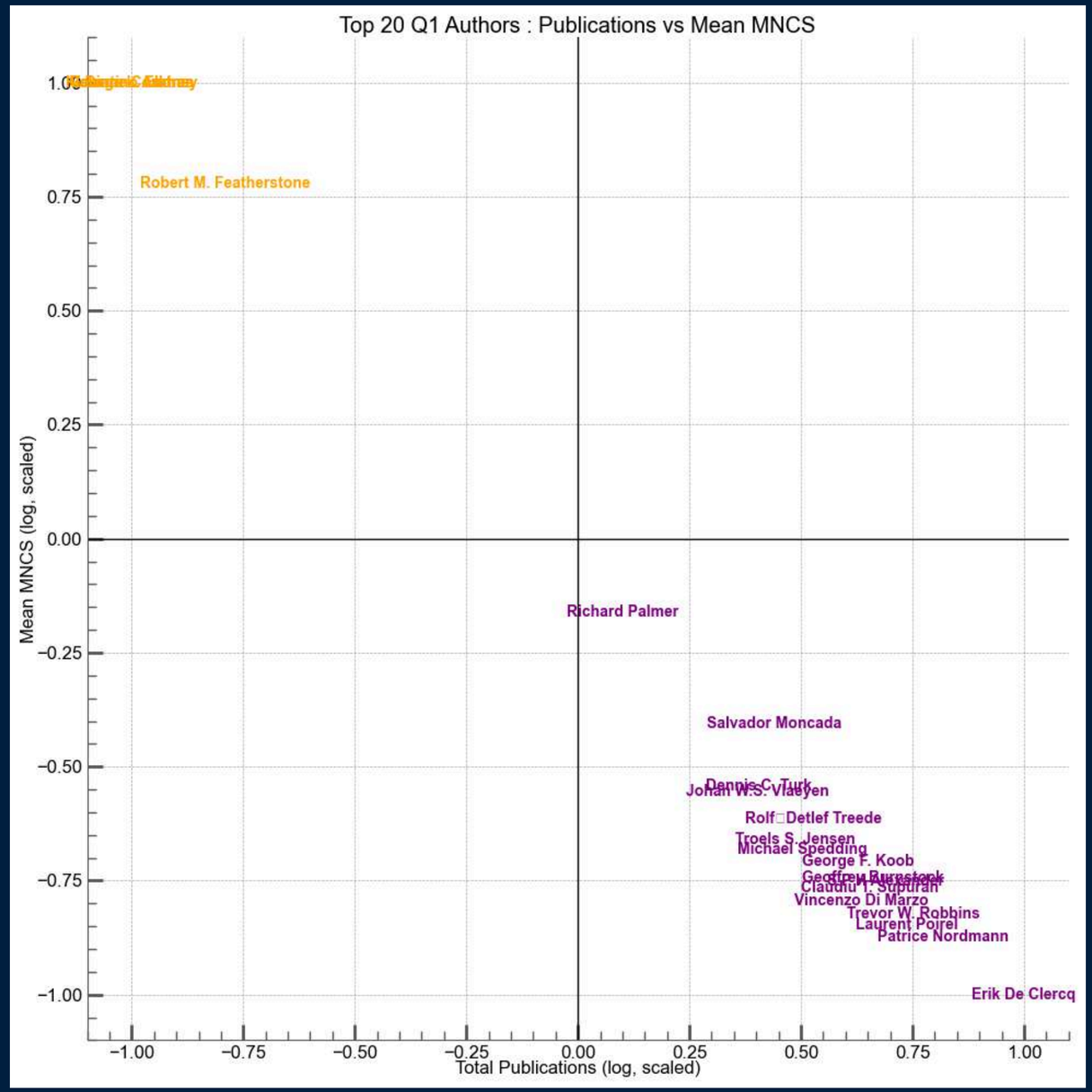
- Authors with more publications tend to have a lower MNCS compared to other top authors
- High MNCS despite many publications (SPH Alexander, Eamonn Kelly)
- Average MNCS despite fewer publications (Vicenzo Di Marzo, Roger G. Pertwee)



Authors

Top Authors - Q1 Publications vs Citations

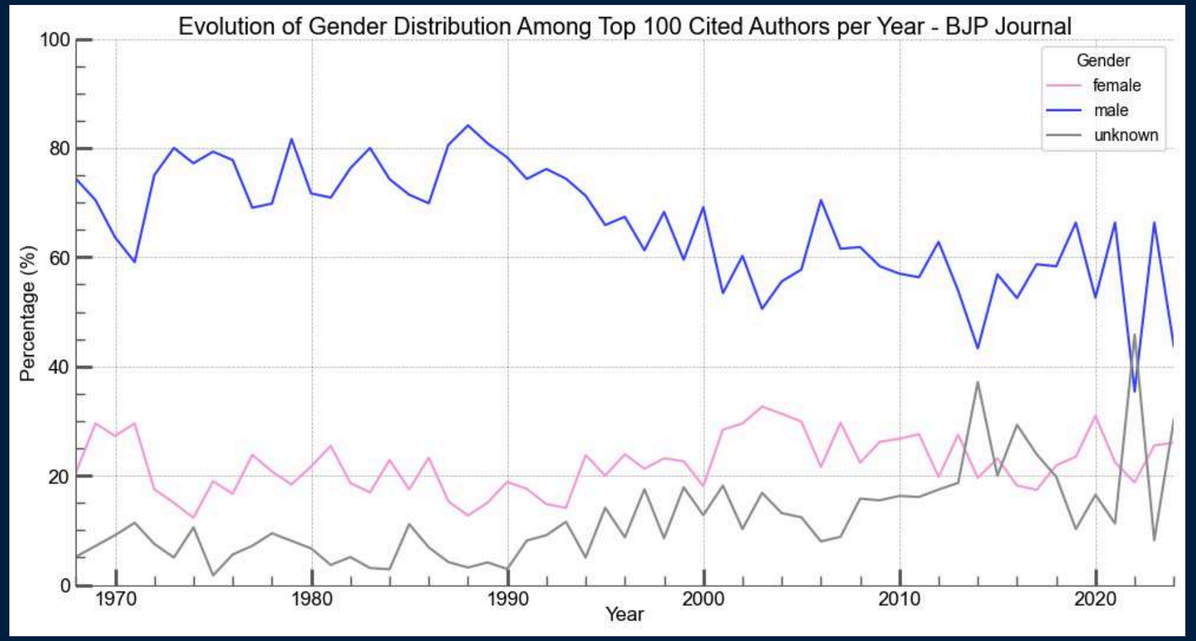
- Authors with more publications tend to have a lower MNCS compared to other top authors

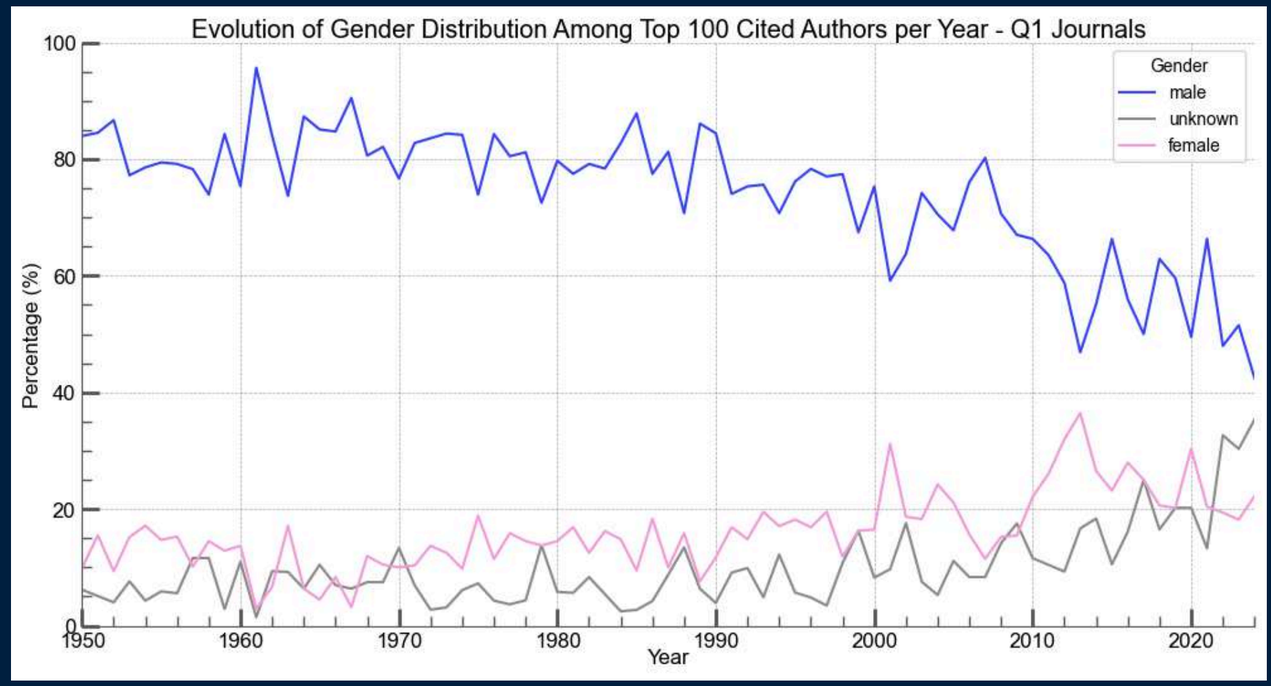


HSDSLab
Human & Social Data Science

Authors

Gender Distribution - Most cited Authors
Gender Detector Python Library



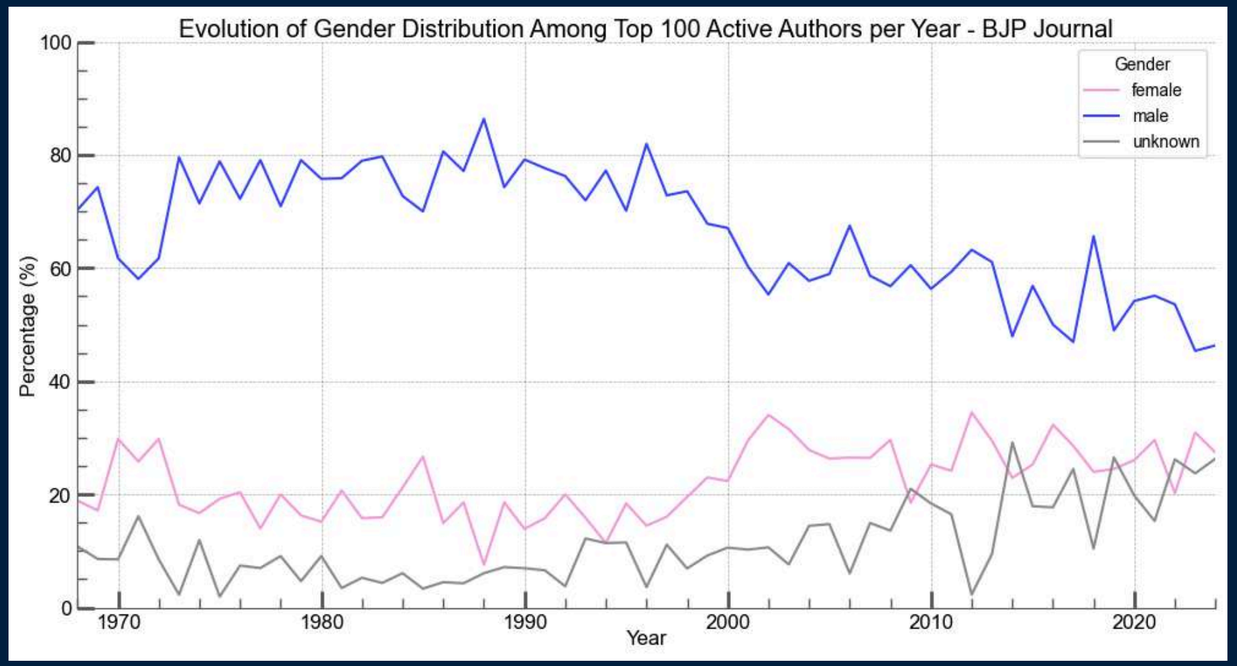


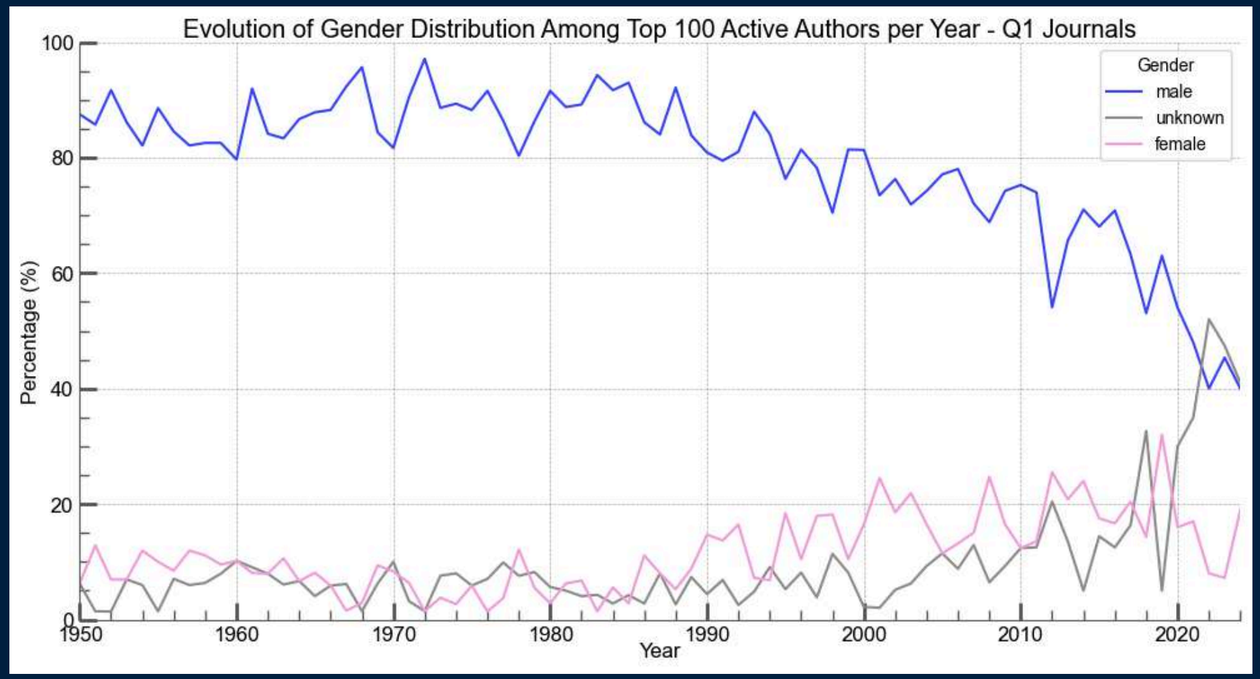
- Predicts the likely gender of a first name.
- Uses a first-name based database to classify names as male, female, or unknown.
- Simple tool for gender analysis in datasets, not perfect.

HSDSLab
Human & Social Data Science

Authors

Gender Distribution - Most A*ctive** Authors
Gender Detector Python Library



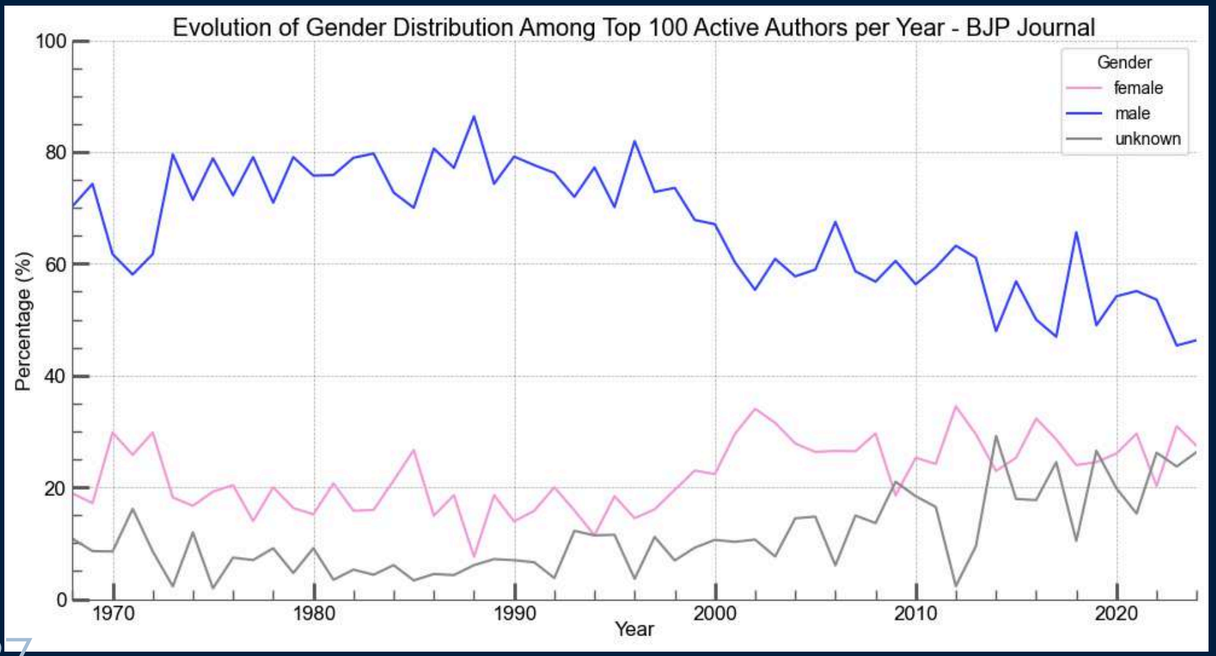


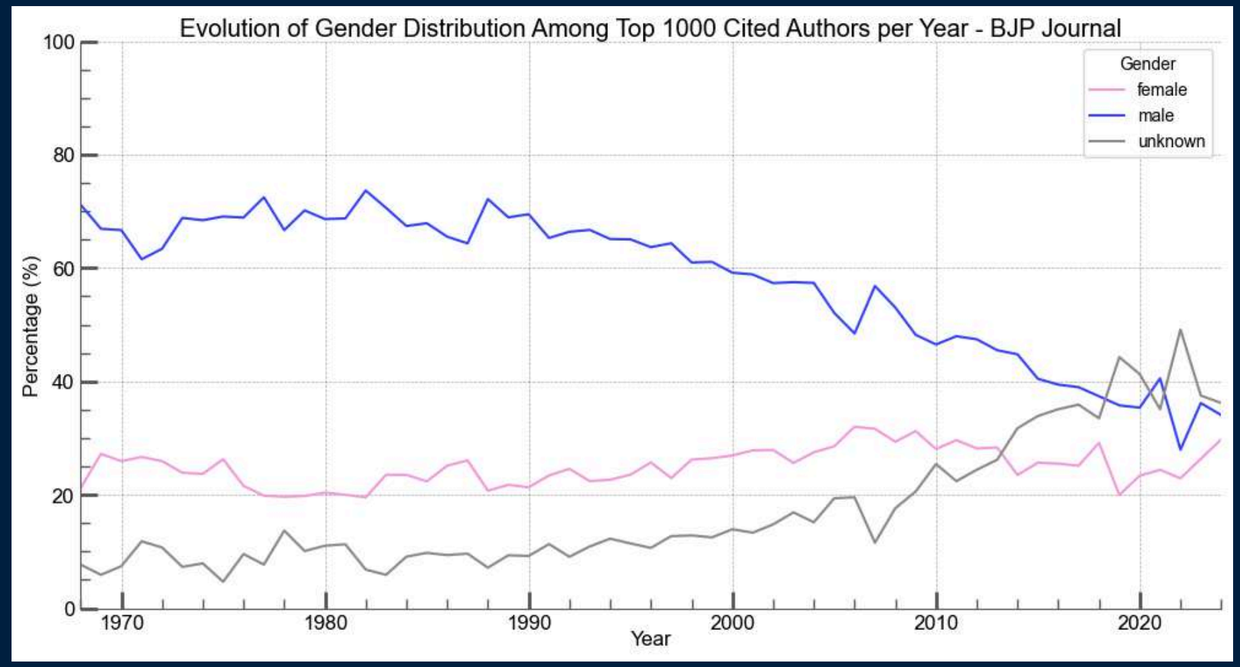
**Active* = With most publications in the year

HSDSLab
Human & Social Data Science

Authors

Gender Distribution - *General* Overview
Gender Detector Python Library





- For technical reasons, *General* means top 1000 cited authors

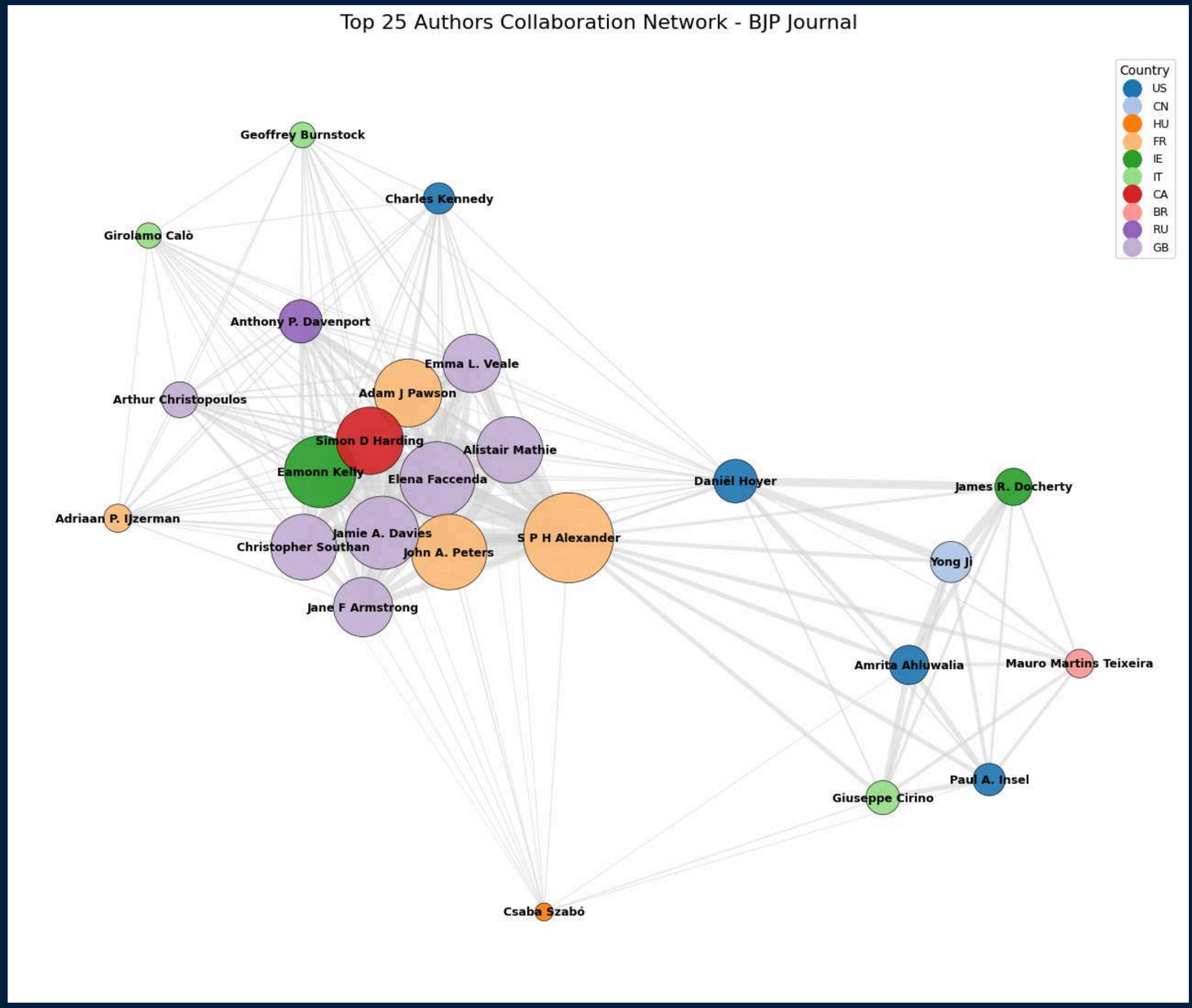
HSDSLab
Human & Social Data Science

Authors

Collaboration - Top 25 authors - BJP

- Selected the top 25 authors by total number of collaborations.
- Collaboration = two authors co-authoring the same article; counted once per article.
- Diverse countries and groups, with small majority of british authors

The author country is based on the country of their main (most frequent) institution. Node sizes are based on degree



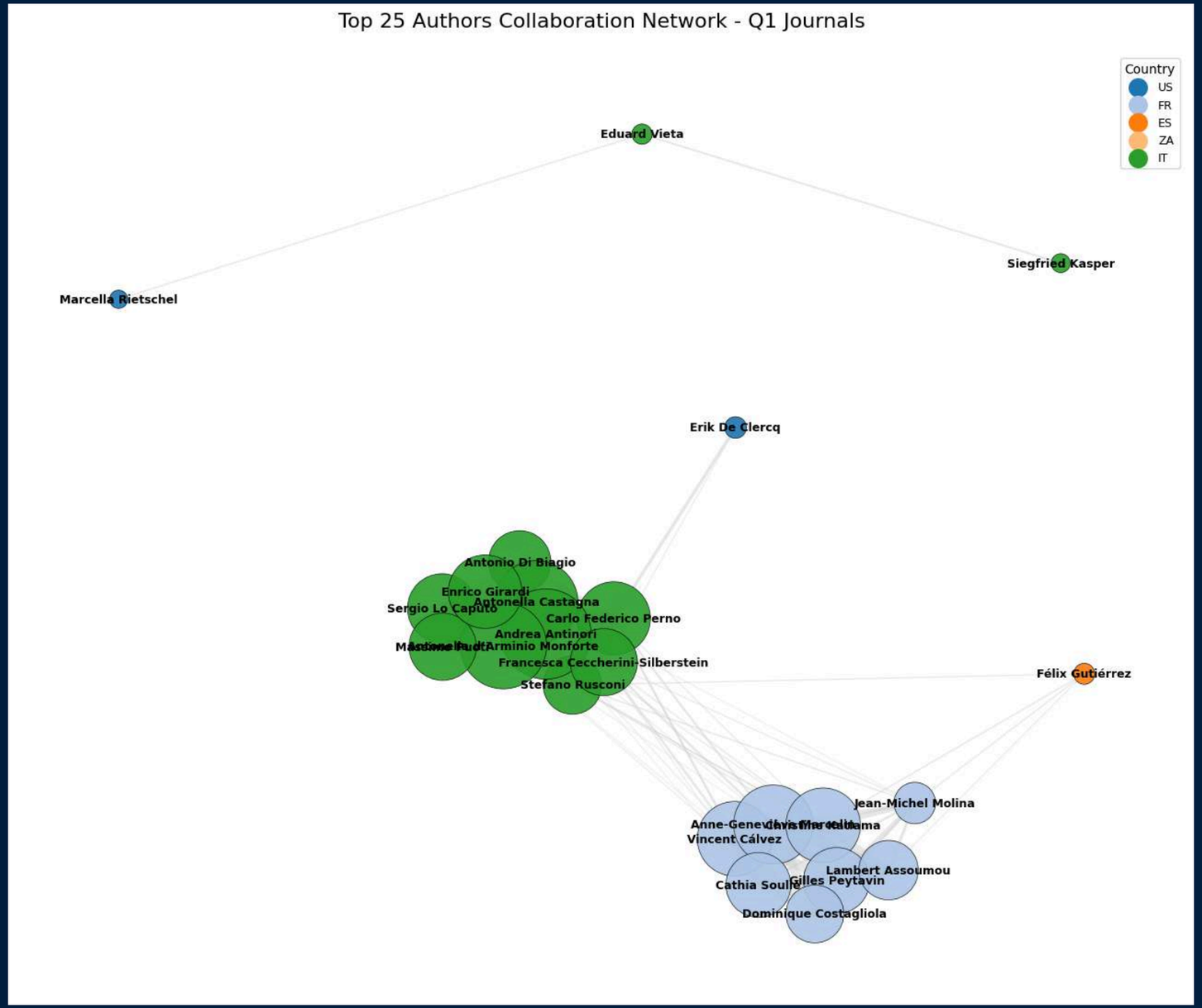
HSDSLab
Human & Social Data Science

Authors

Collaboration - Top 25 authors - Q1

- Selected the top 25 authors by total number of collaborations.
- Collaboration = two authors co-authoring the same article; counted once per article.
- Two large groups: Italian authors and French authors collaborate a lot within their own group

The author country is based on the country of their main (most frequent) institution. Node sizes are based on degree



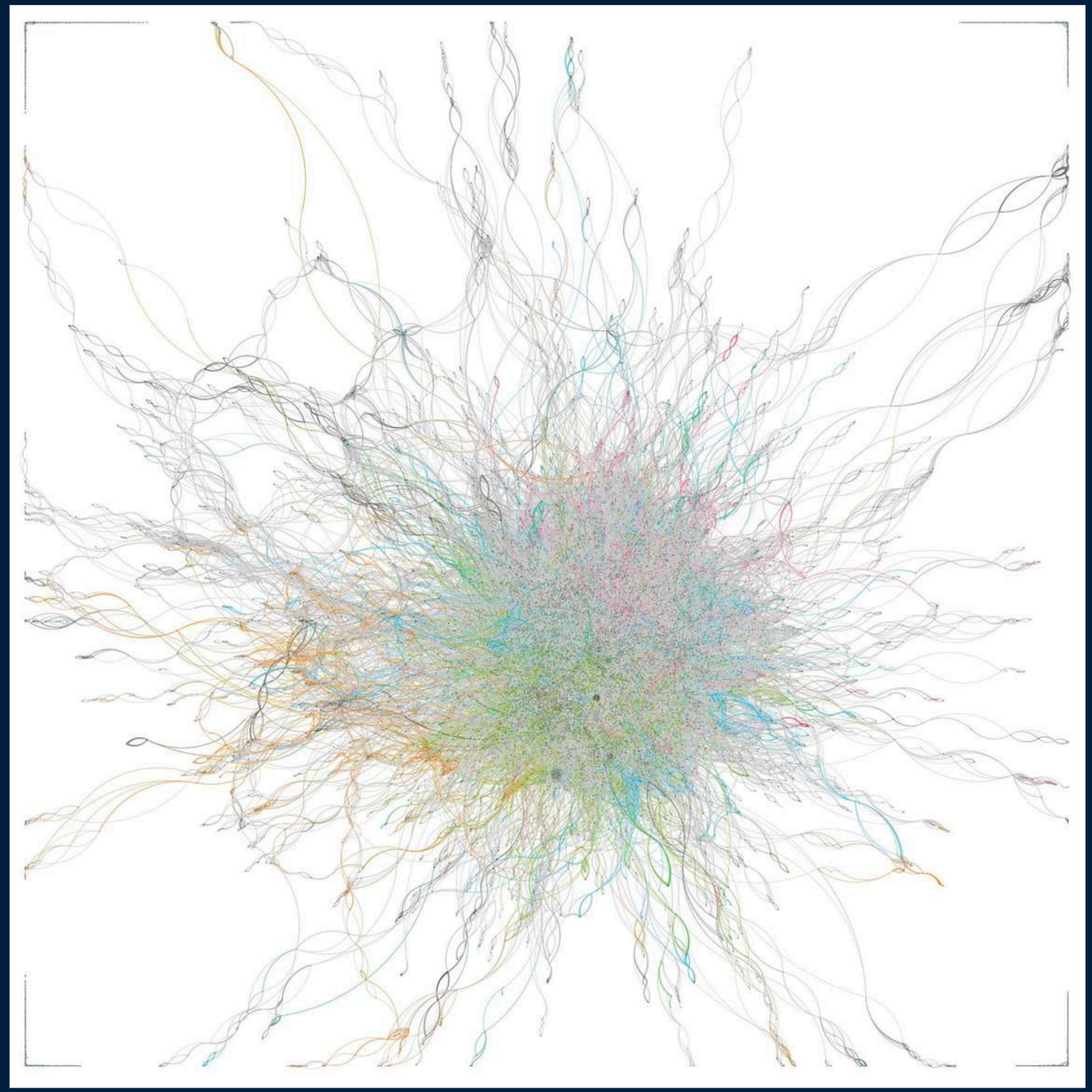
Authors

Collaboration - BJP Co-Authorship Network

- Made with Gephi

- Colored by authors's countries :

GB	(4.41%)
US	(3.44%)
CN	(2.35%)
JP	(1.89%)
DE	(1.49%)
FR	(1.3%)



HSDSLab
Human & Social Data Science

Authors

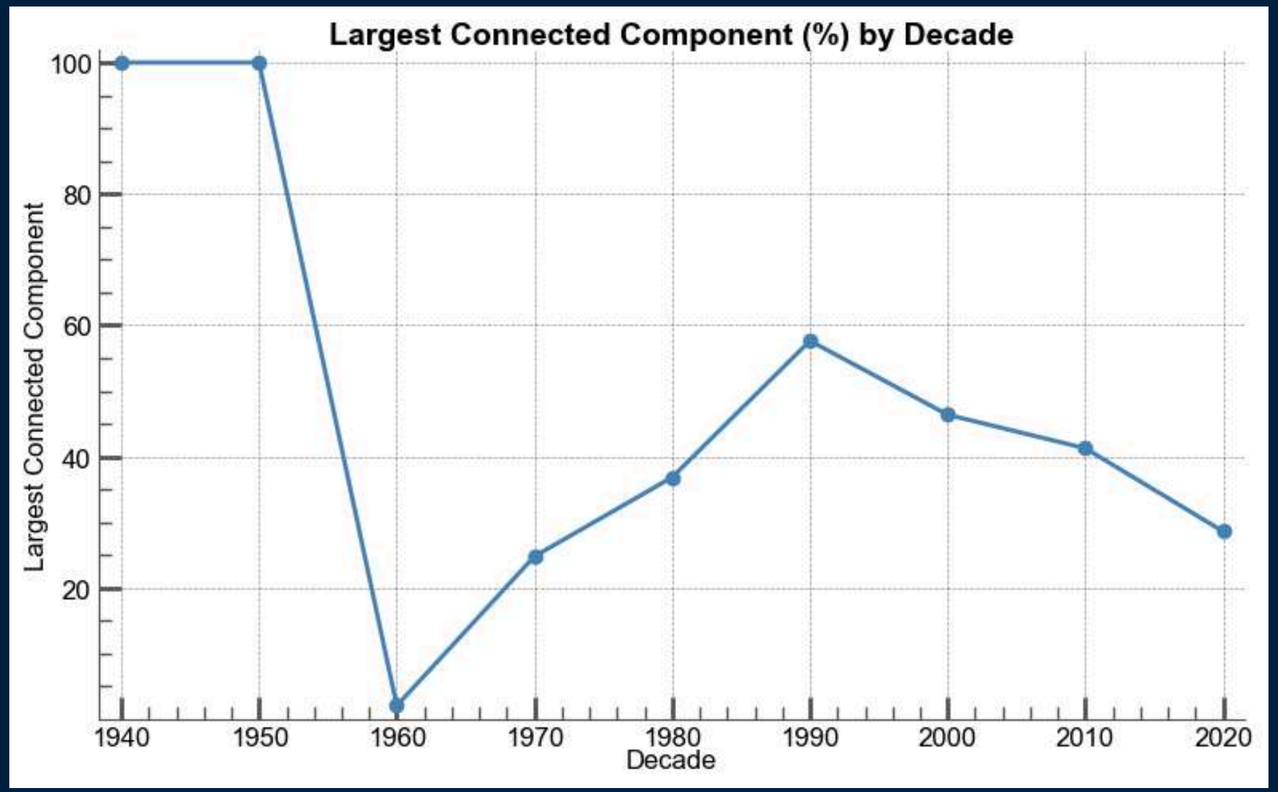
Co-Authorship Network - BJP LCC %

- Strong evolution from 1960s to 1990s (Peak at around 58%)
- However, a small decline up to the 2020s (around 30%)

→ For the BJP journal, connectedness of scientists increases from the 60s until the 90s but then decreases

LCC = Largest Connected Component

The largest group of authors who are all connected through collaborations.



HSDSLab
Human & Social Data Science

Authors

Co-Authorship Network - Q1 LCC %

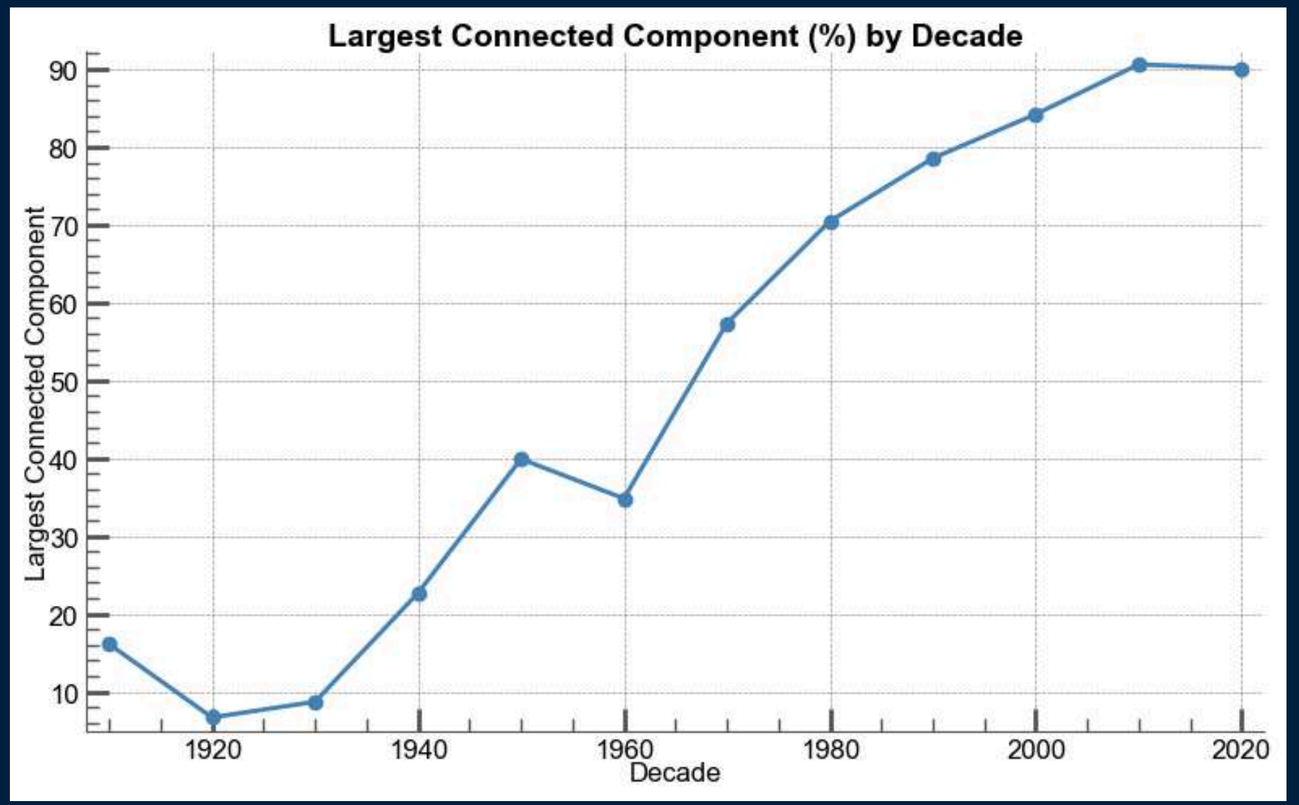
LCC = Largest Connected Component

The largest group of authors who are all connected through collaborations.

- Strong evolution through the decades

(Peak at around 91% in the 2010s)

→ For Q1 journals, connectedness of scientists increases over the time



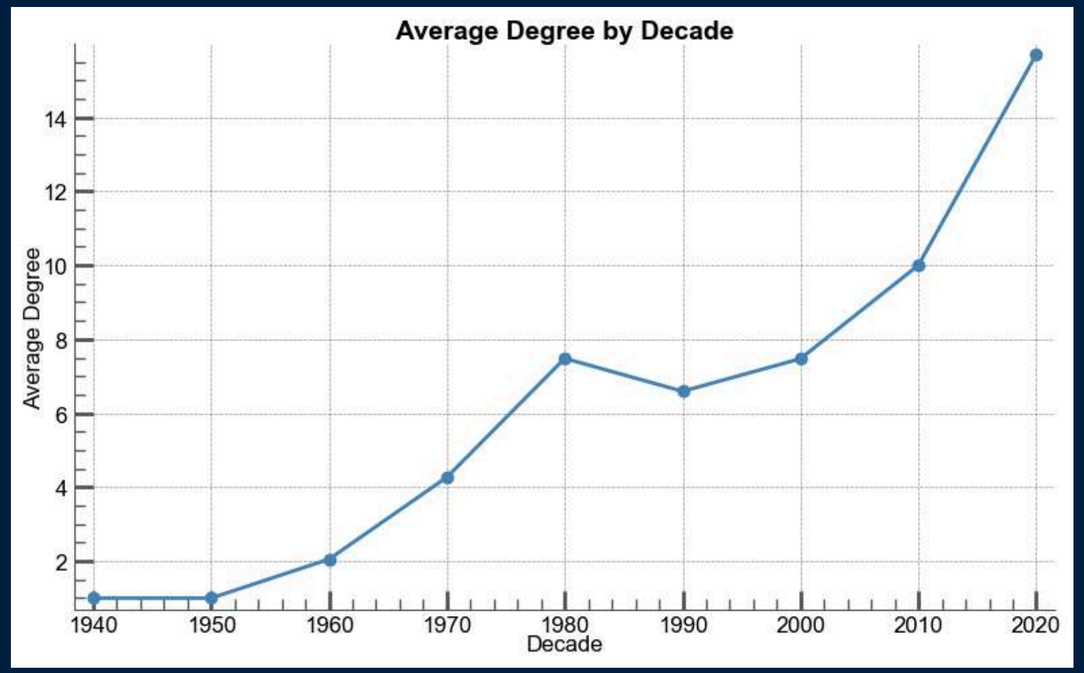
HSDSLab
Human & Social Data Science

Authors

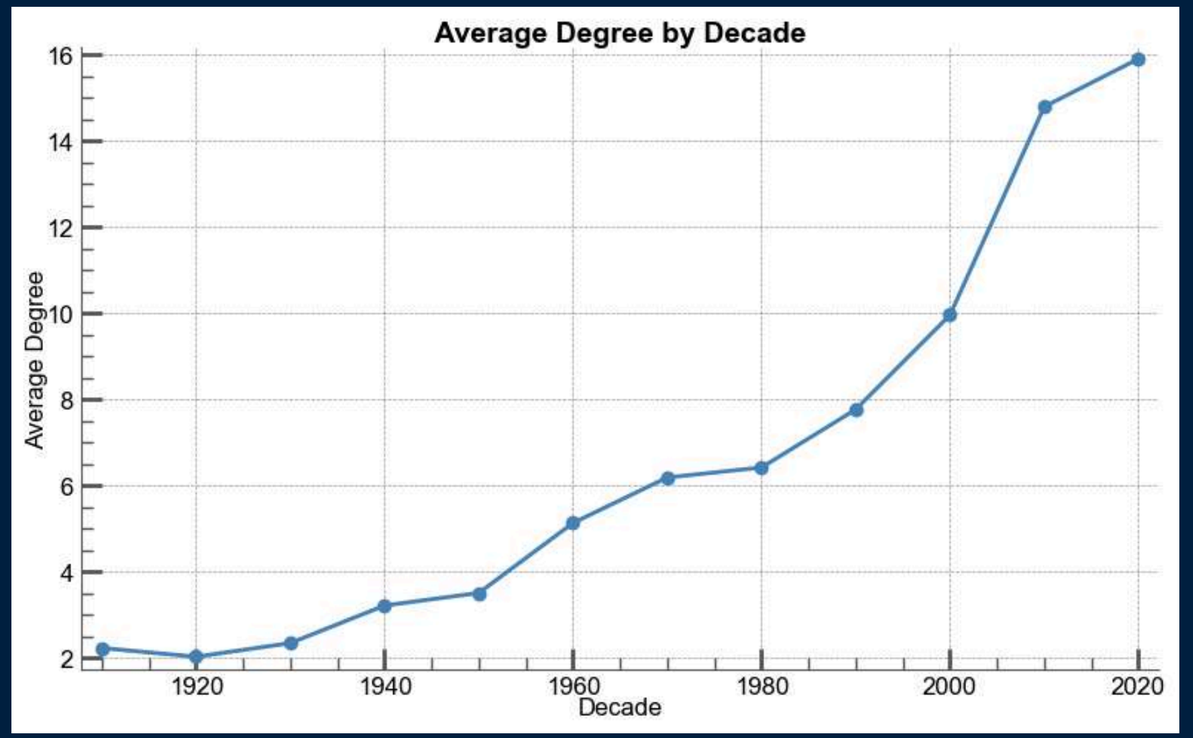
Co-Authorship Network - Average Degree

Average Degree :

The average number of collaborators each author has.



Avg ˚ for BJP



Avg ˚ for Q1 Journals

HSDSLab
Human & Social Data Science

Authors

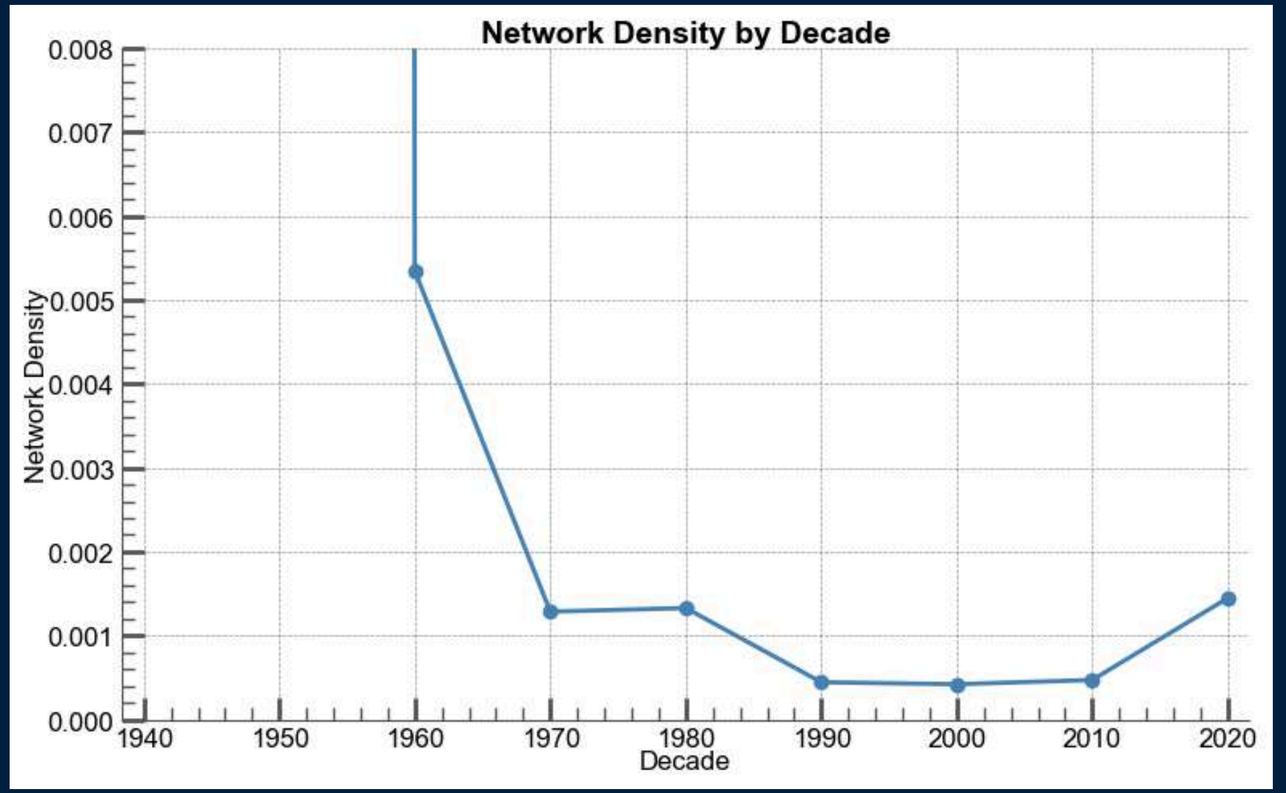
Co-Authorship Network - BJP Network Density

Network Density :

The proportion of all possible collaboration links that actually exist in the network.

- The network density logically decreases given the higher number of authors over the decades

- The 2020s show that groups are more well-connected internally nowadays but weakly connected to each other.



HSDSLab
Human & Social Data Science

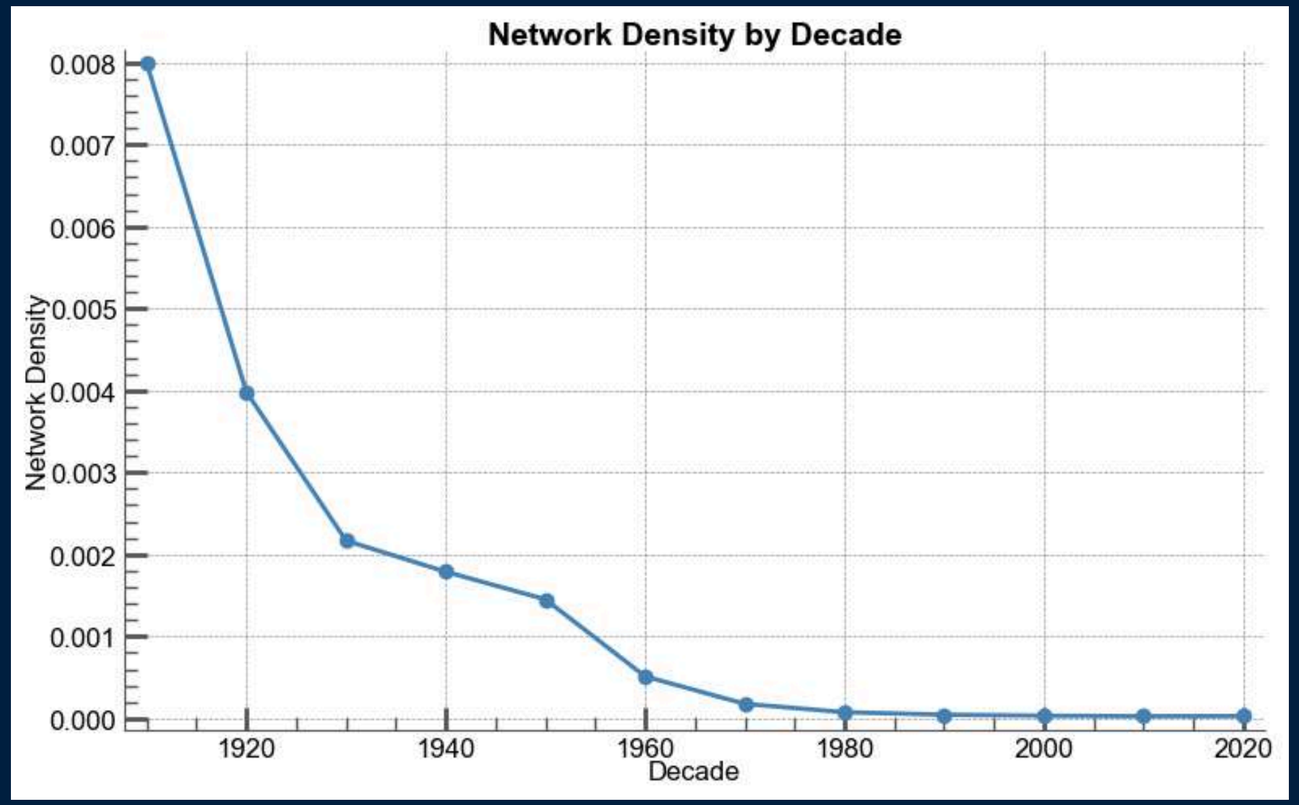
Authors

Co-Authorship Network - Network Density

Network Density :

The proportion of all possible collaboration links that actually exist in the network.

- The network density logically decreases given the higher number of authors over the decades



HSDSLab
Human & Social Data Science

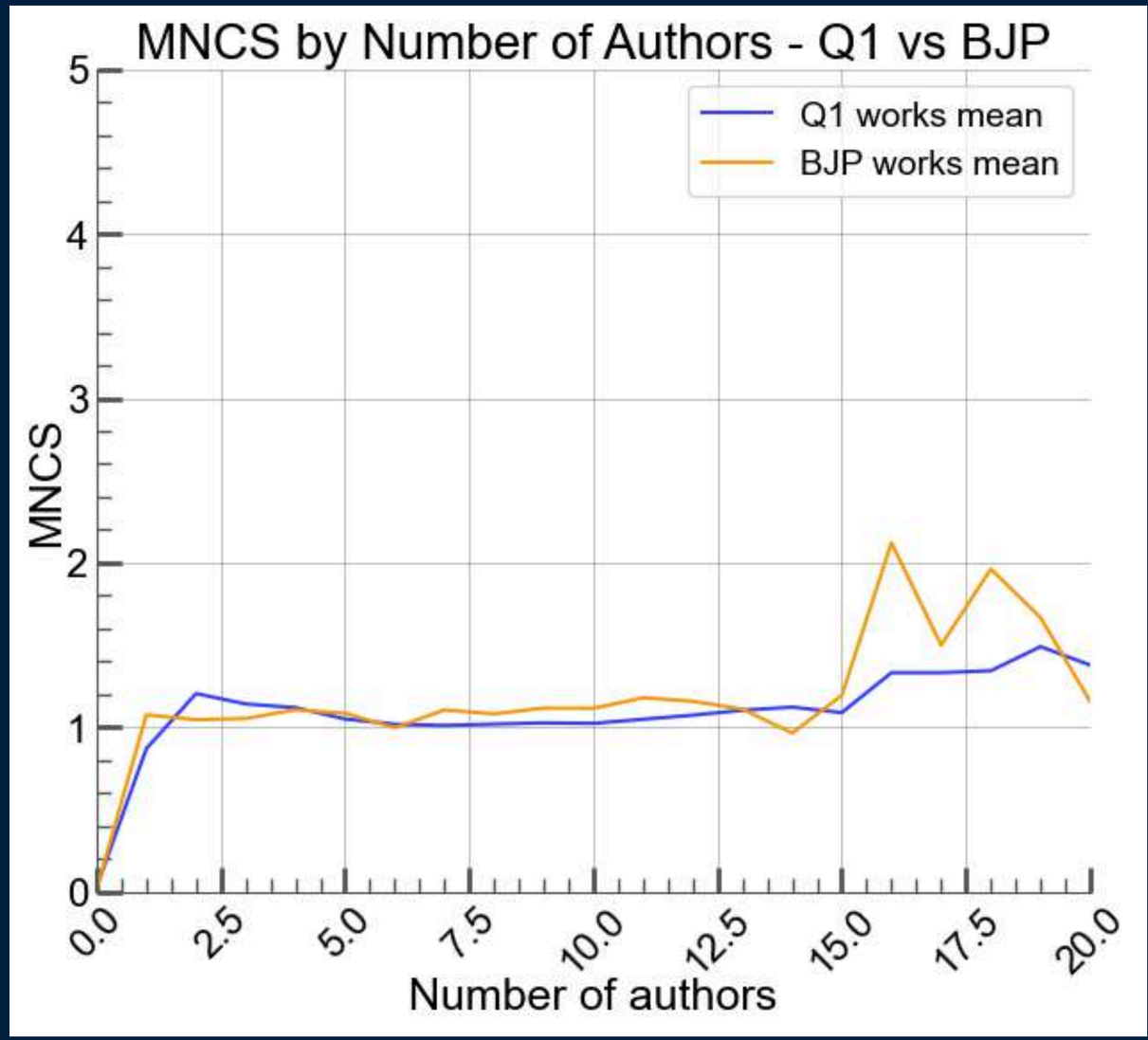
Authors

Related to MNCS

- Number of authors capped at 20 for clear visualization (too much variation otherwise)
- Curve fairly stable, little impact for this range
- Can be tested statistically (all numbers)

```
Pearson: r = 0.045, p-value = 0.000
Spearman: rho = 0.279, p-value = 0.000
```

→ Small but meaningful trend over a wider range



HSDSLab
Human & Social Data Science

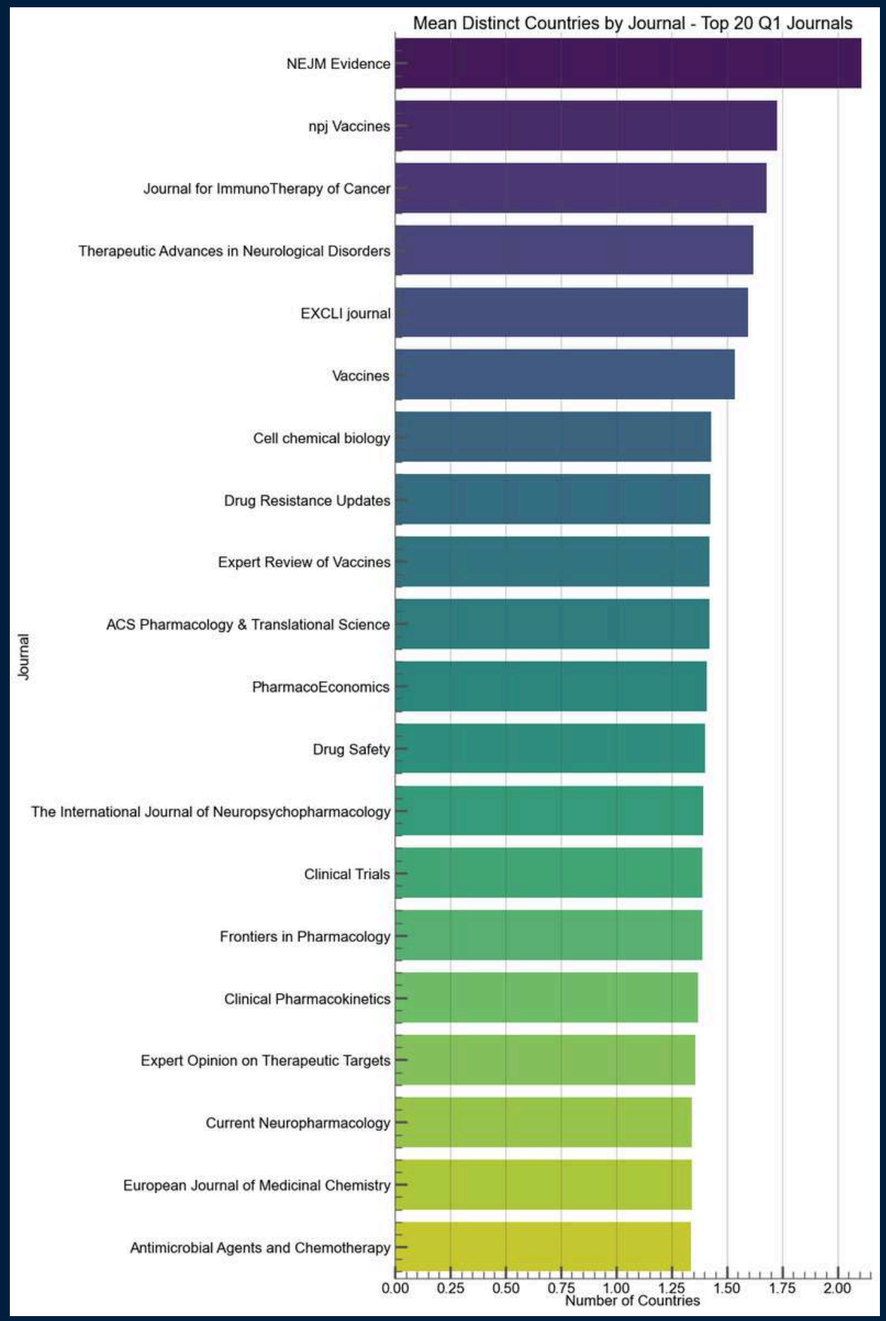
Countries

Countries

Journal Ranking

BJP is ranked **32nd**, with around

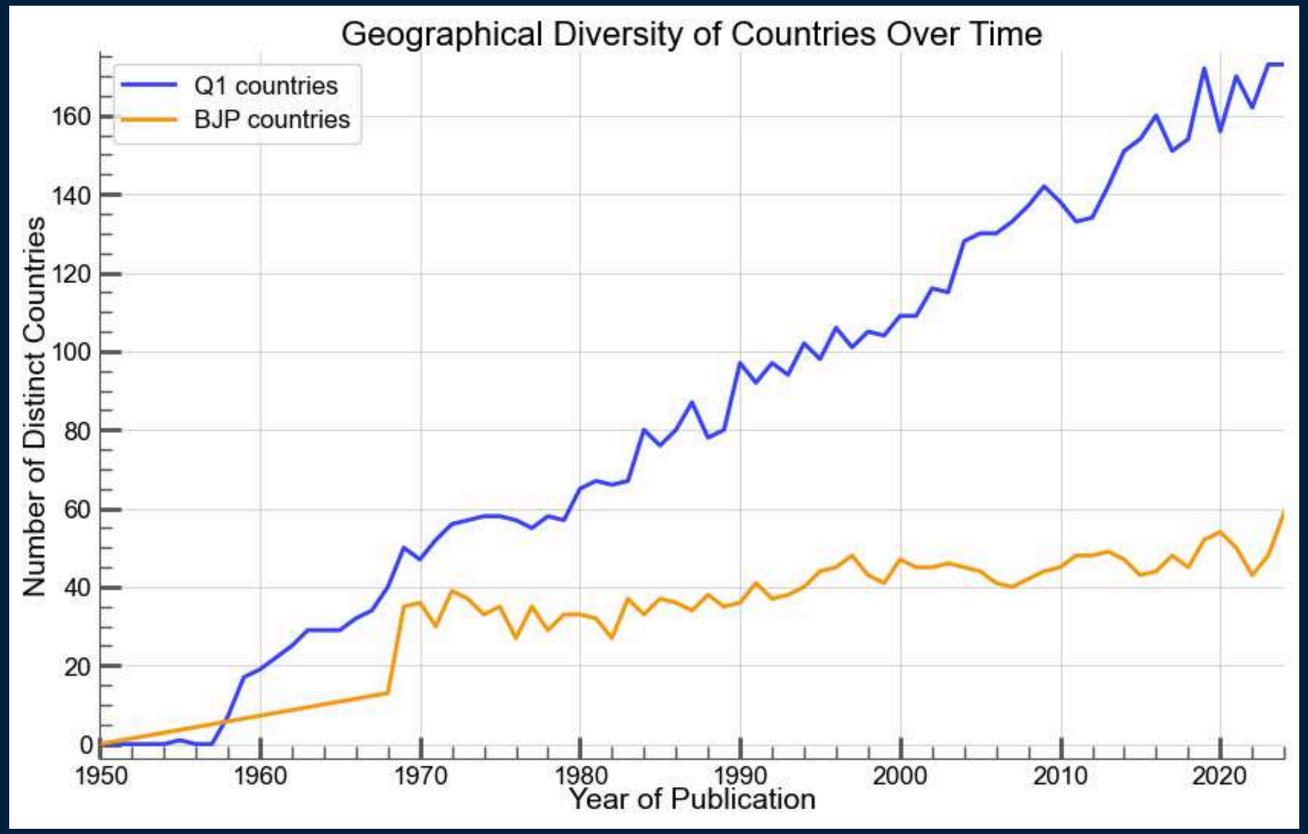
1.29 countries per work.

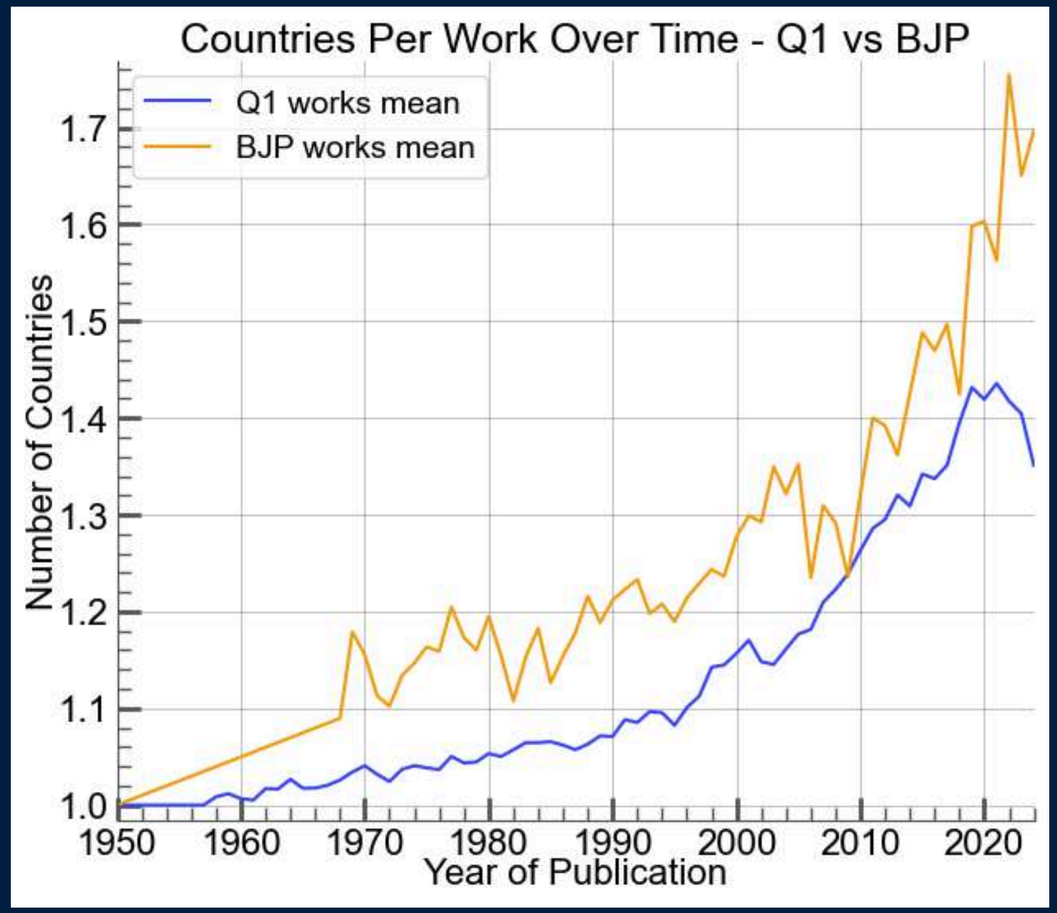


HSDSLab
Human & Social Data Science

Countries

General Overview





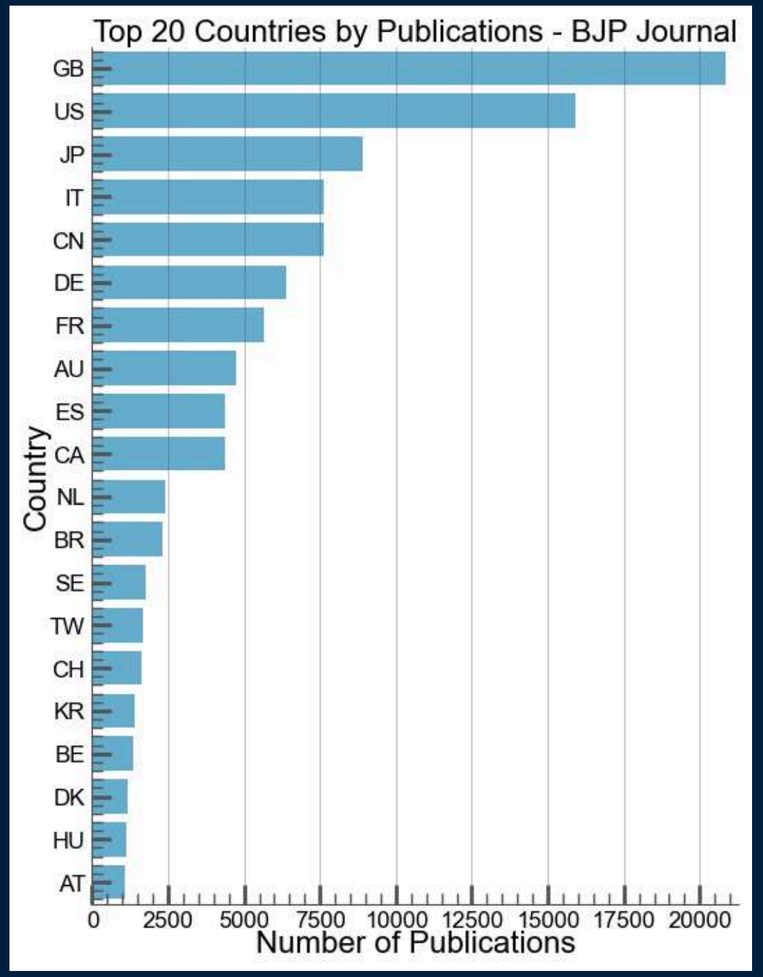
Peak in 2024 for BJP : 59 Countries

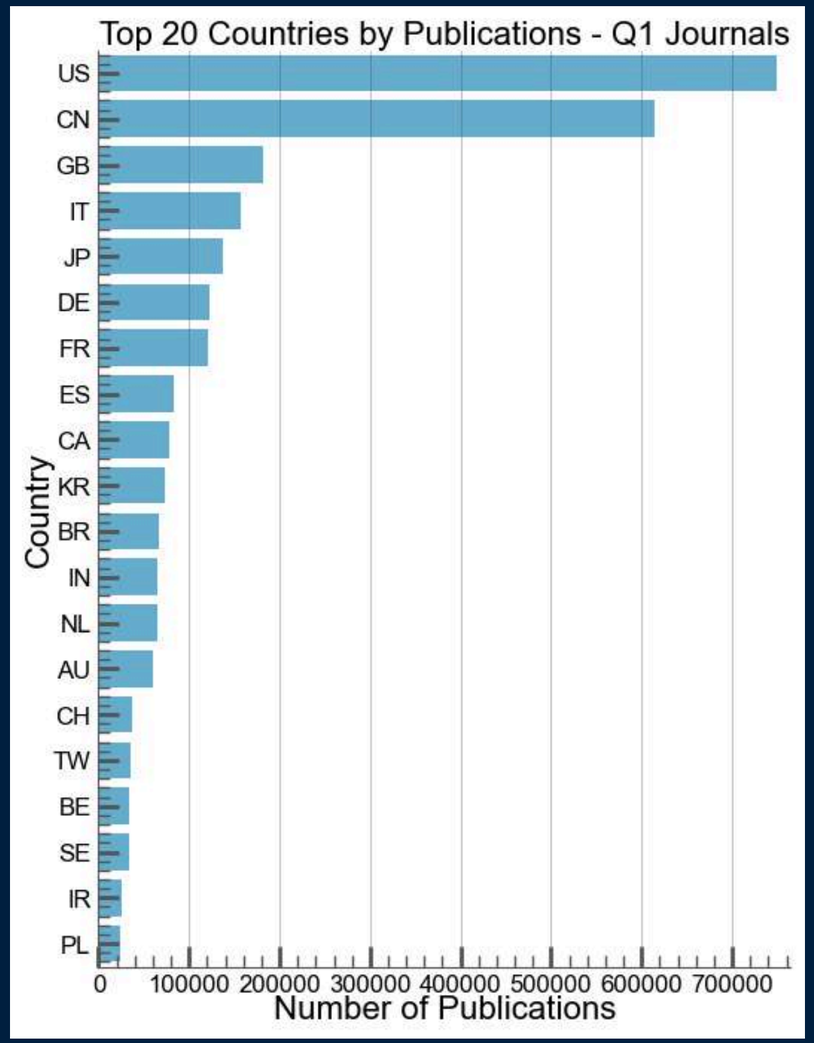
Peak in 2024 for Q1 Journals : 173 Countries

HSDSLab
Human & Social Data Science

Countries

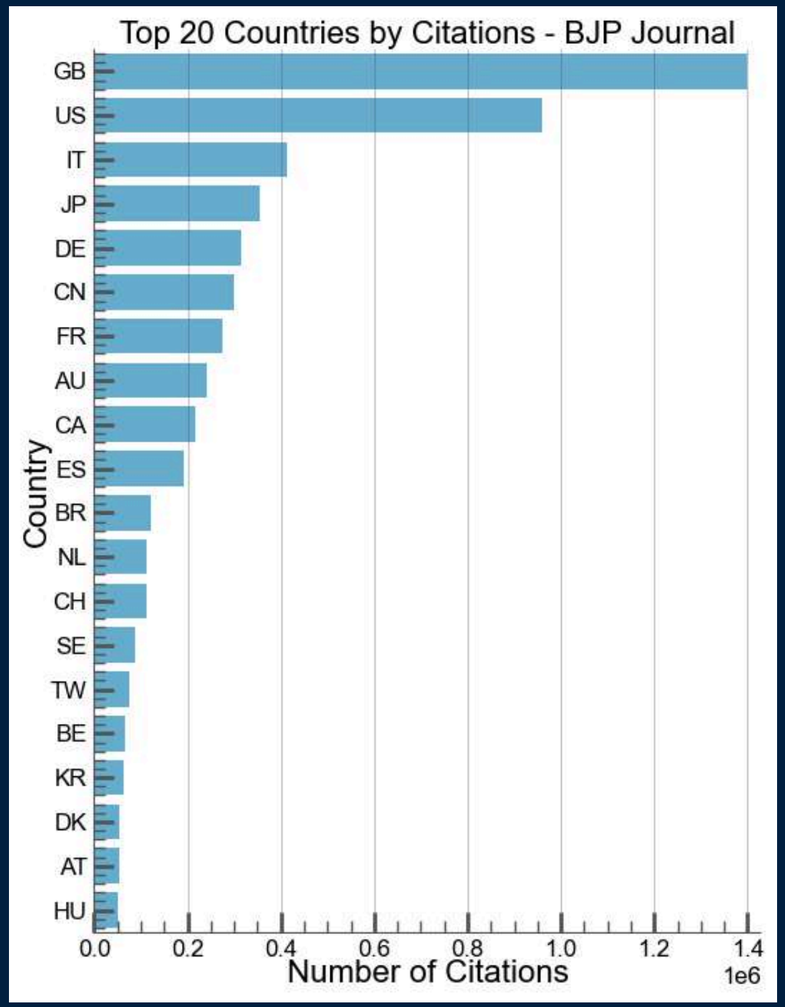
Top Countries - Total Publications

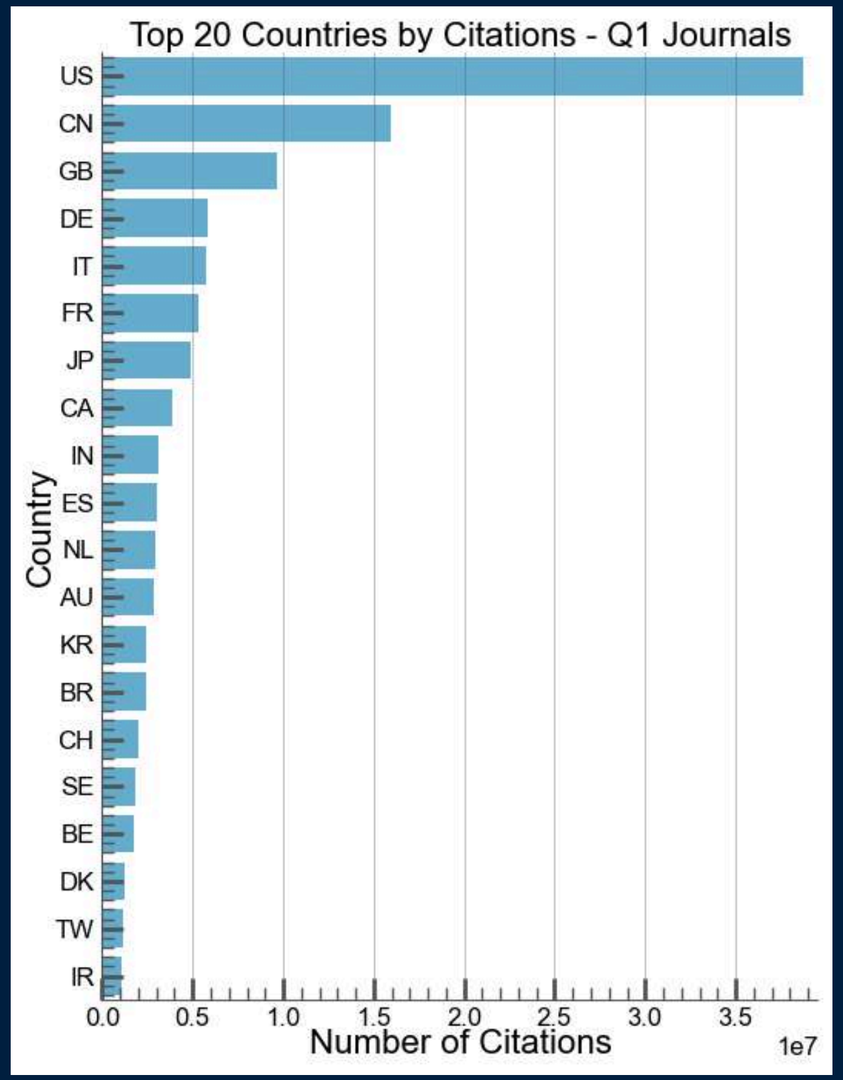




Countries

Top Countries - Total Citations

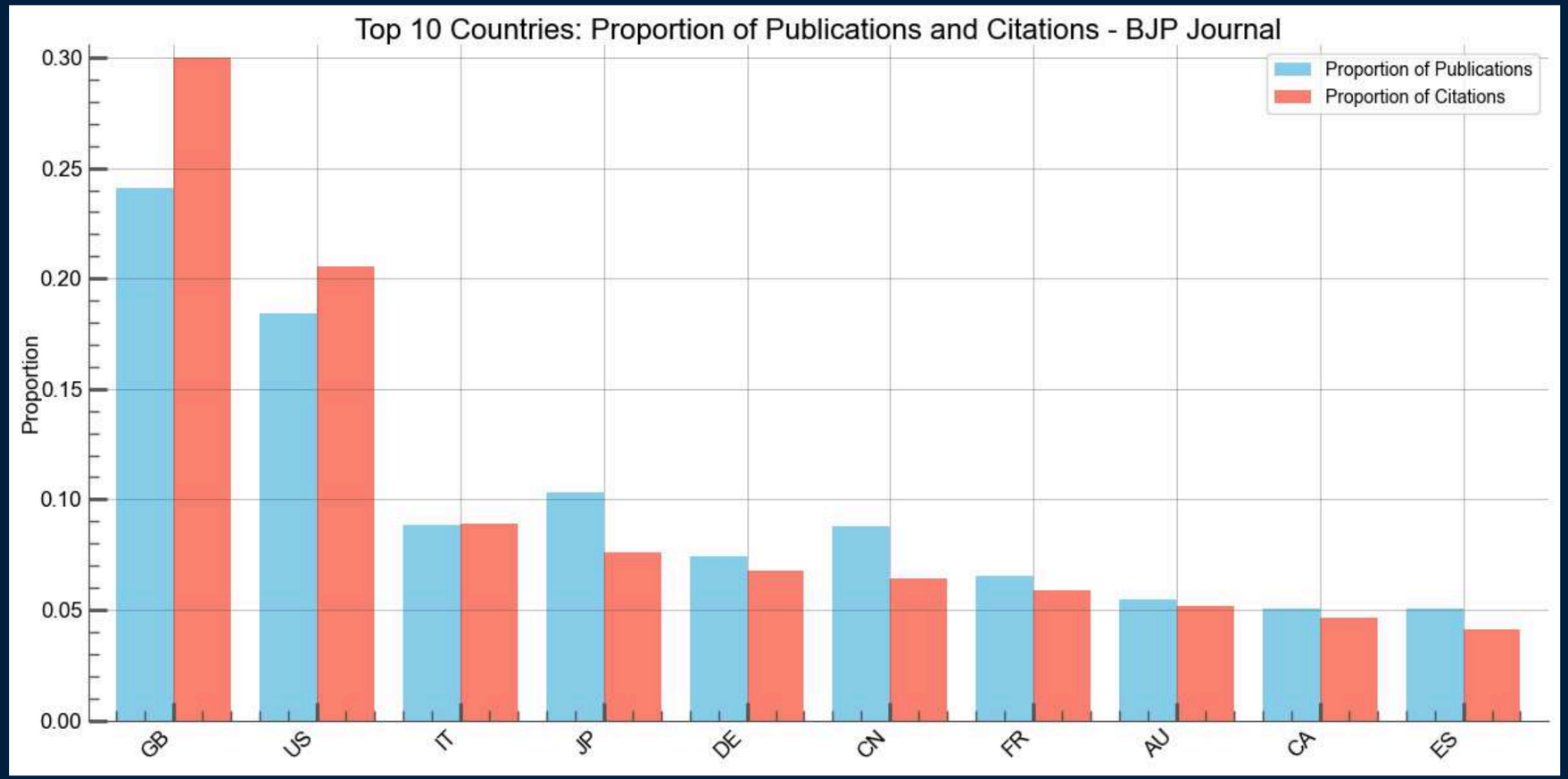




HSDSLab
Human & Social Data Science

Countries

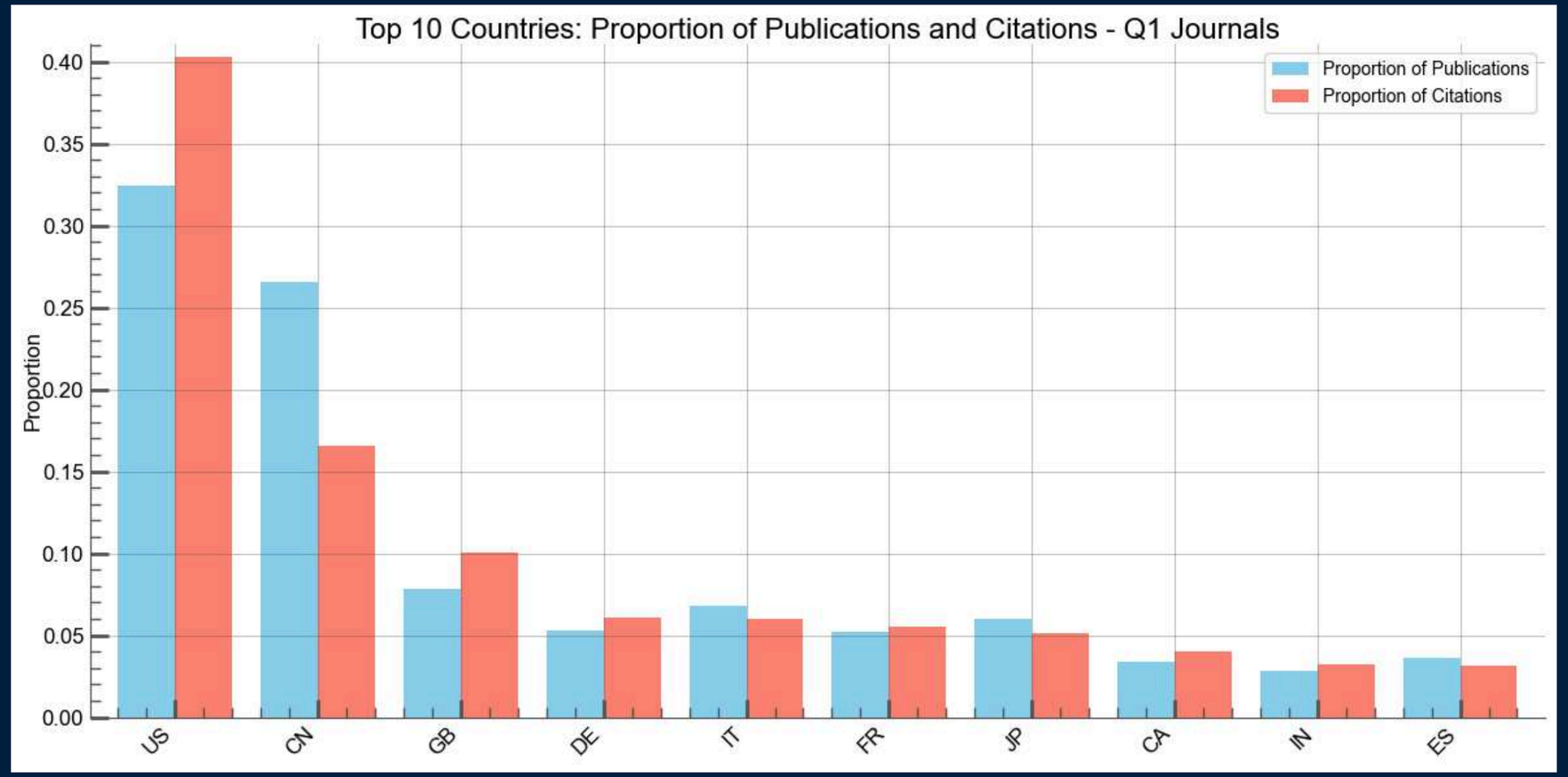
Top Countries - Citations/Publications - BJP



HSDSLab
Human & Social Data Science

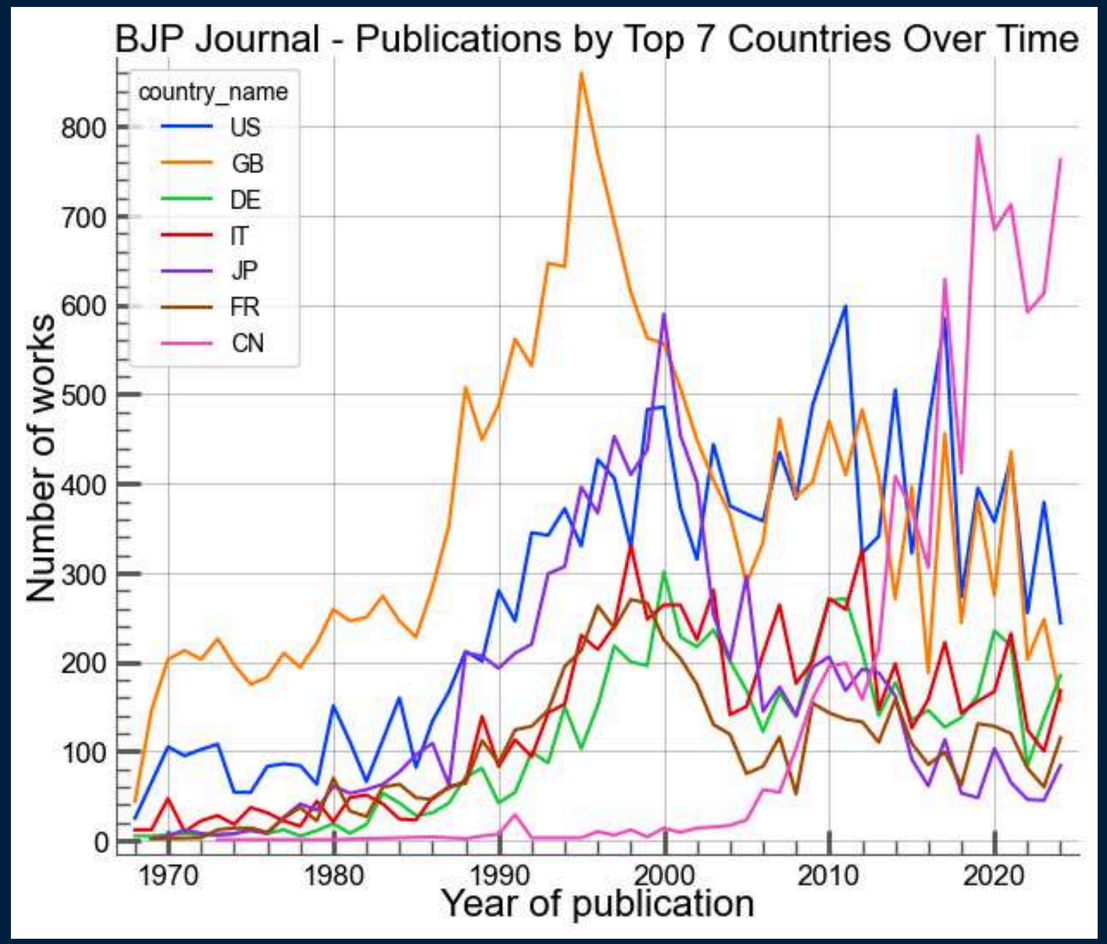
Countries

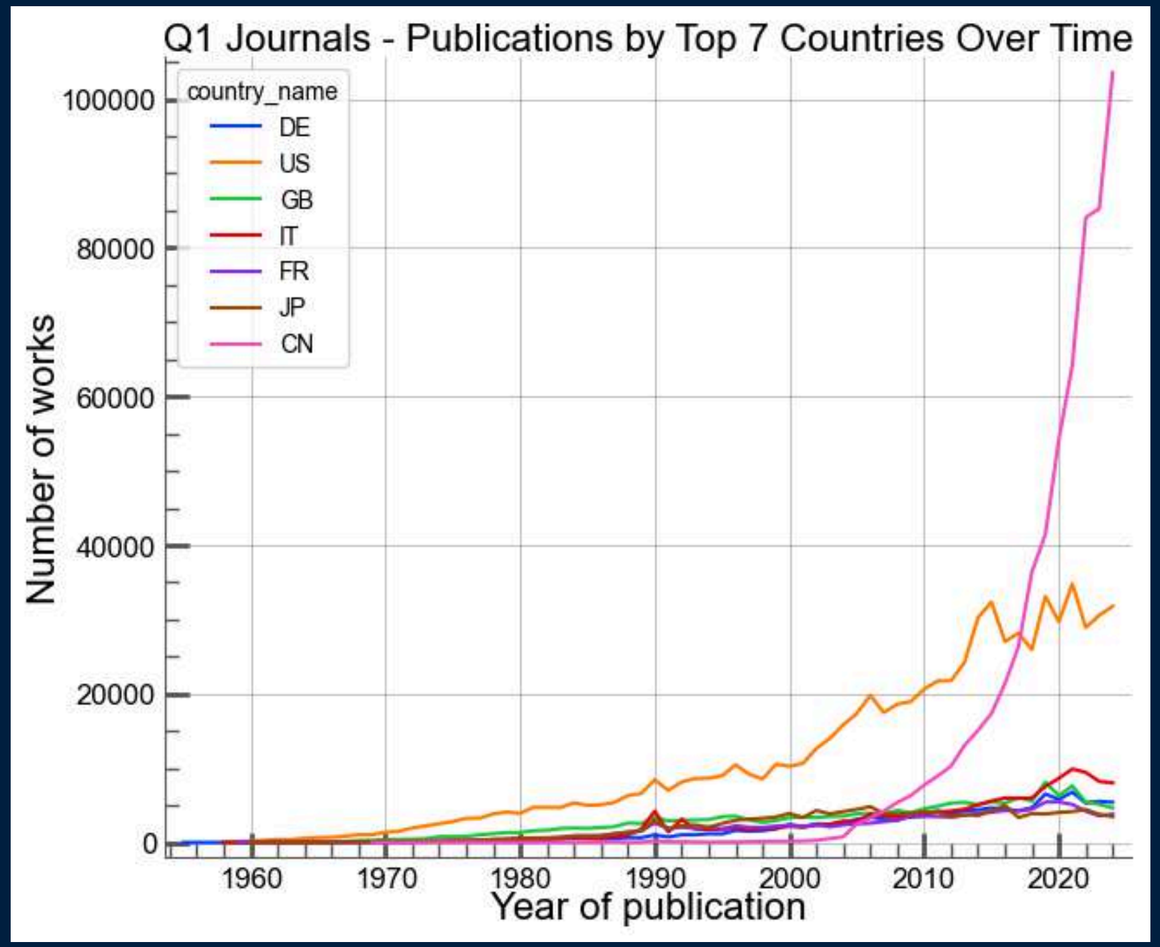
Top Countries - Citations/Publications - Q1 Journals



Countries

Top Countries - Evolution of Publications



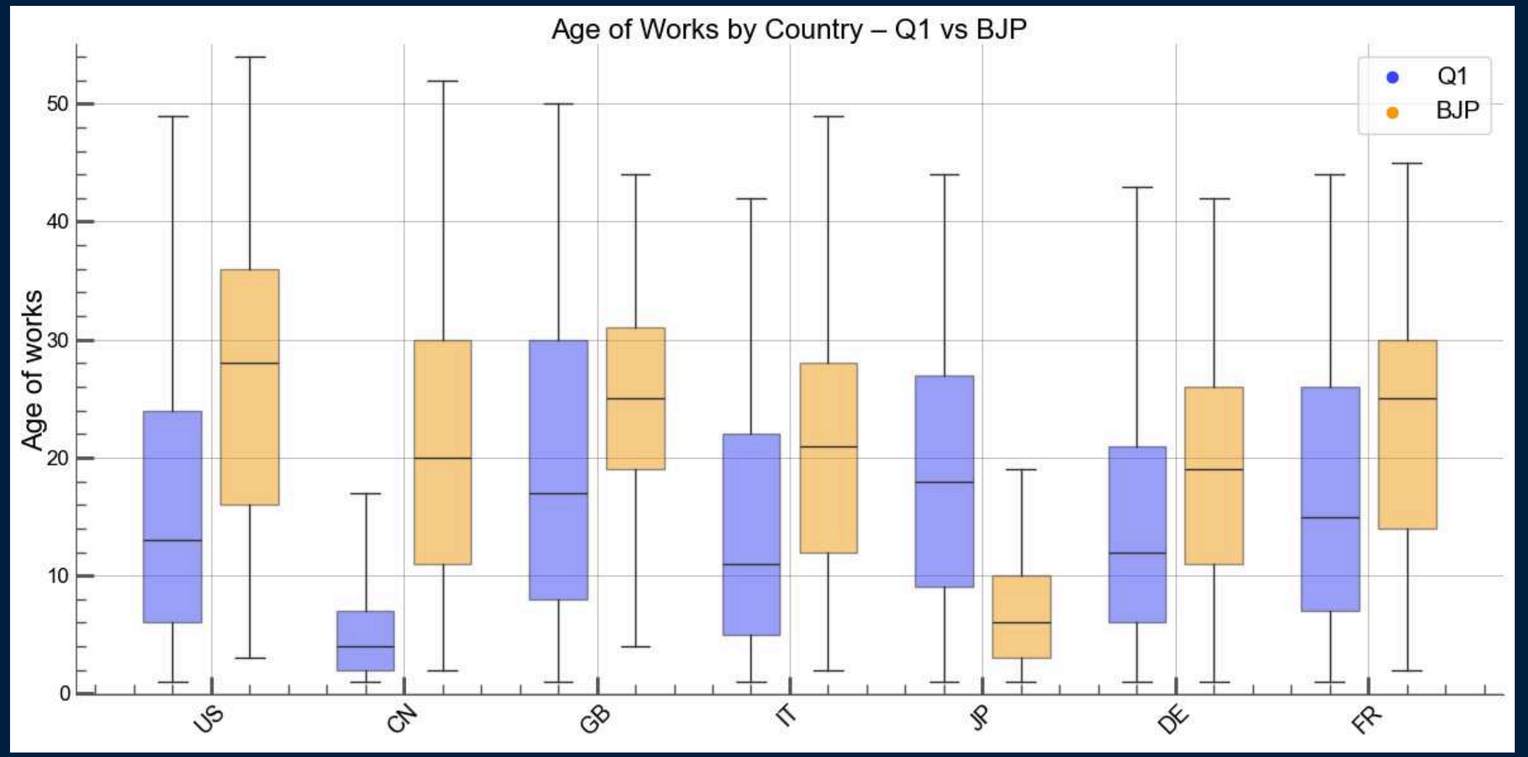


HSDSLab
Human & Social Data Science

Countries

Top Countries - Mean Age of Works

- Strong rise of China
- Marked decline for the UK, especially in BJP.
- Consistently strong presence of the United States

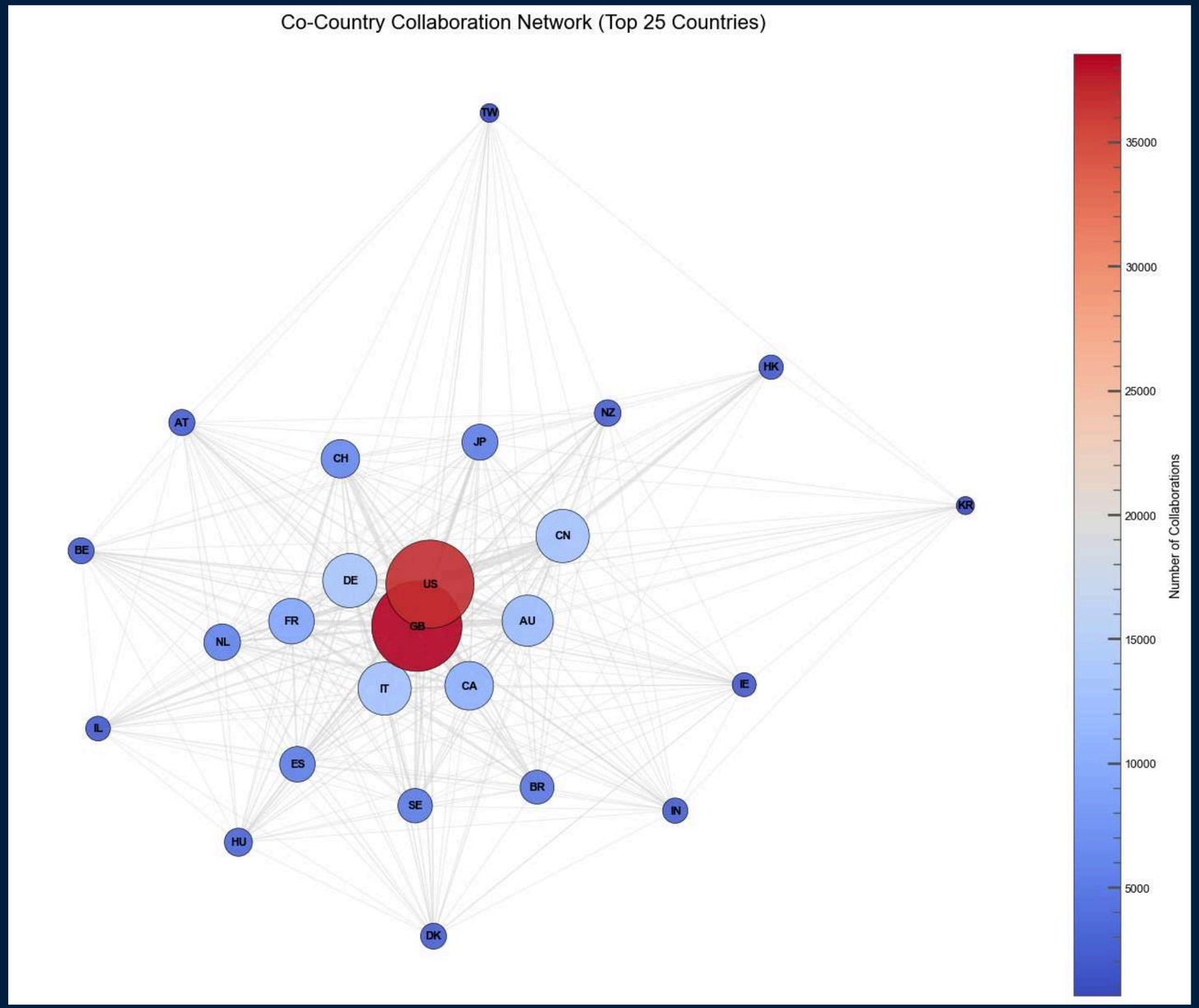


HSDSLab
Human & Social Data Science

Countries

Top Countries - BJP Collaboration

- Selected the top 25 countries by total number of collaborations.
- Collaboration = two countries co-authoring the same article; counted once per article.
- Countries from same continent seem closer, high influence from US/GB.

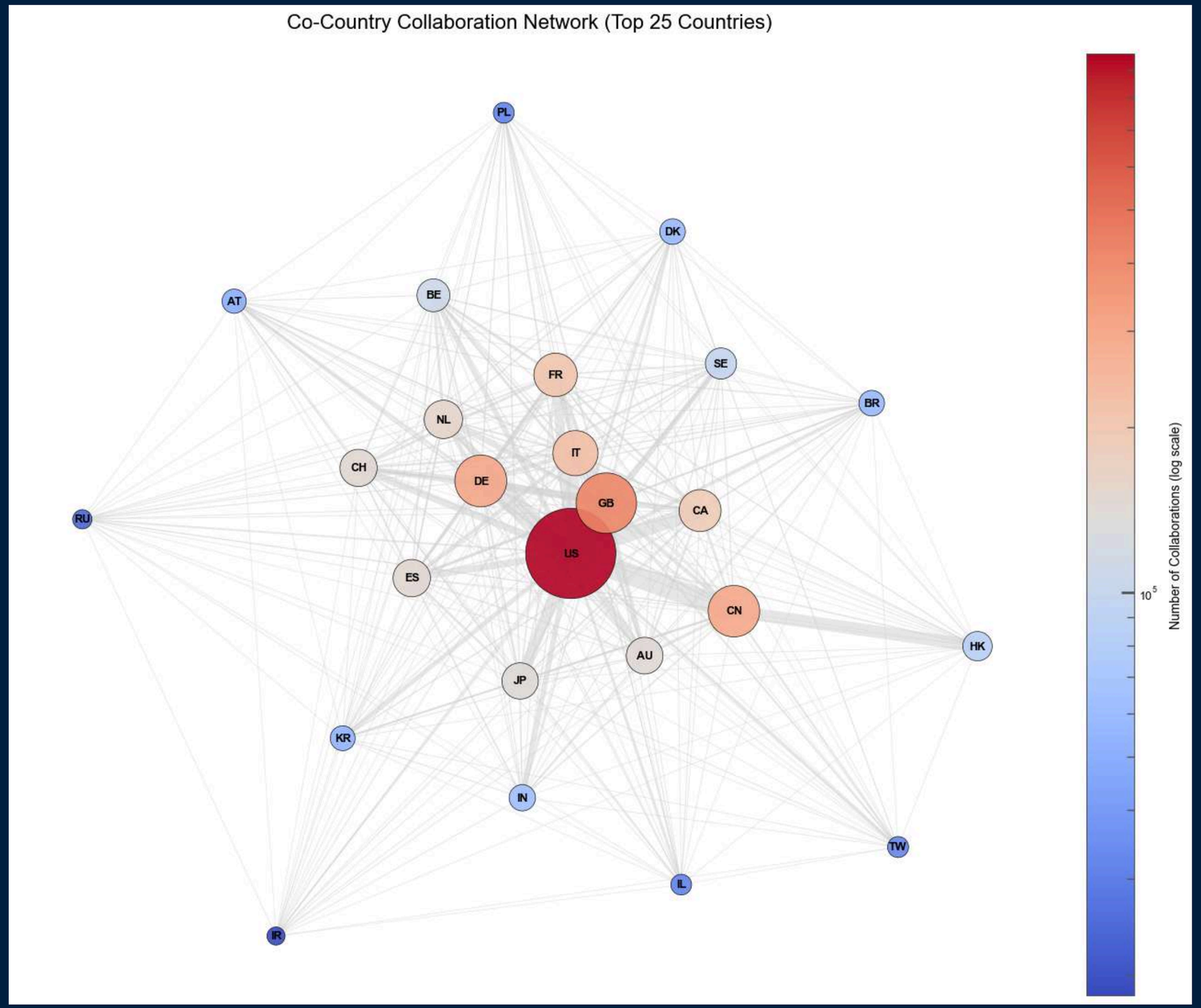


HSDSLab
Human & Social Data Science

Countries

Top Countries - Q1 Collaboration

- Selected the top 25 countries by total number of collaborations.
- Collaboration = two countries co-authoring the same article; counted once per article.
- Countries from same continent seem closer, “underwelming” influence from China ?

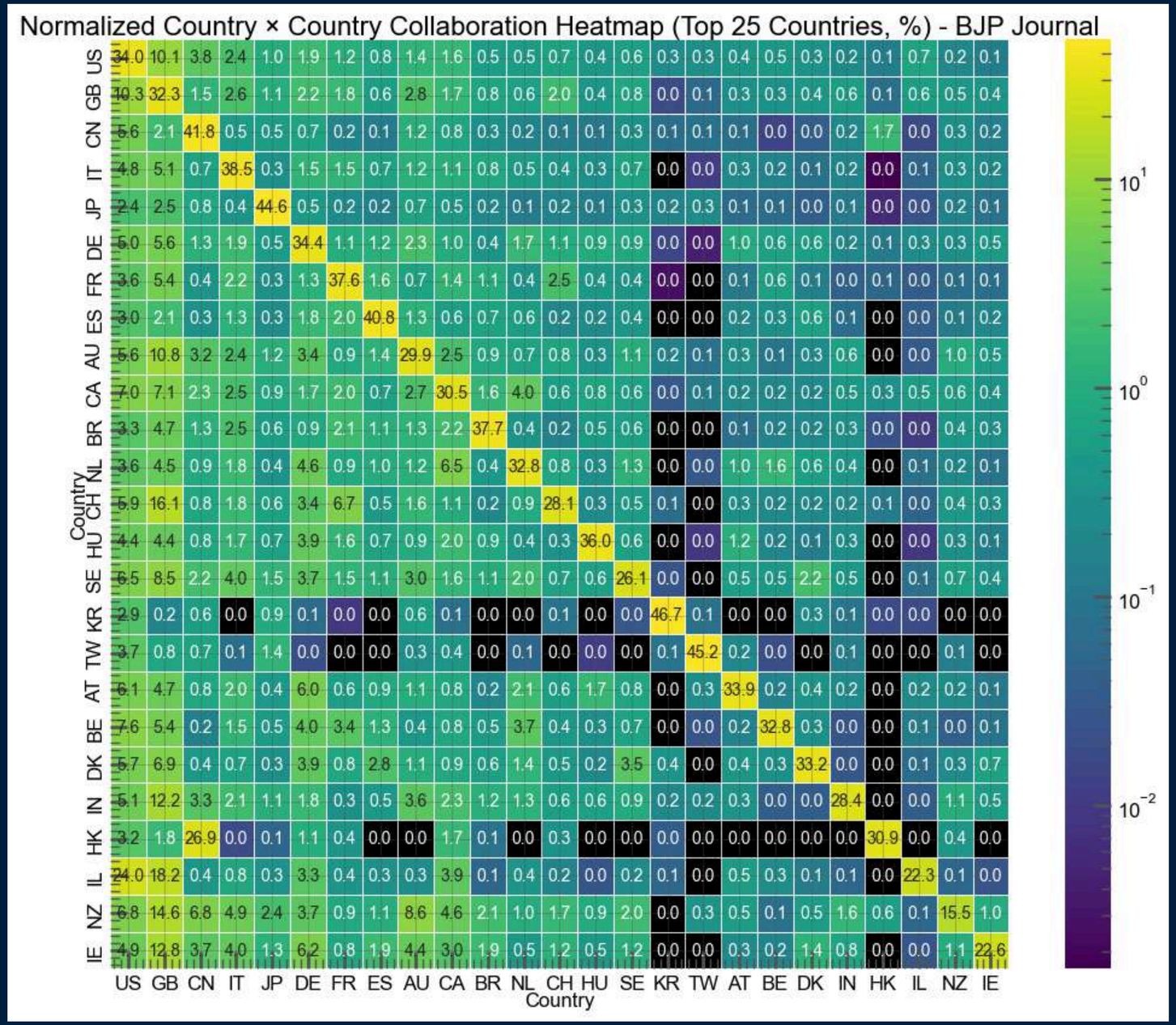


Countries

Top Countries - BJP Collaboration

- Rows = each country's collaboration profile → How this country distributes its collaborations across other countries. (Each number : % of the country's collabs)

- Columns = each country's importance as a partner → How much other countries collaborate with this country.

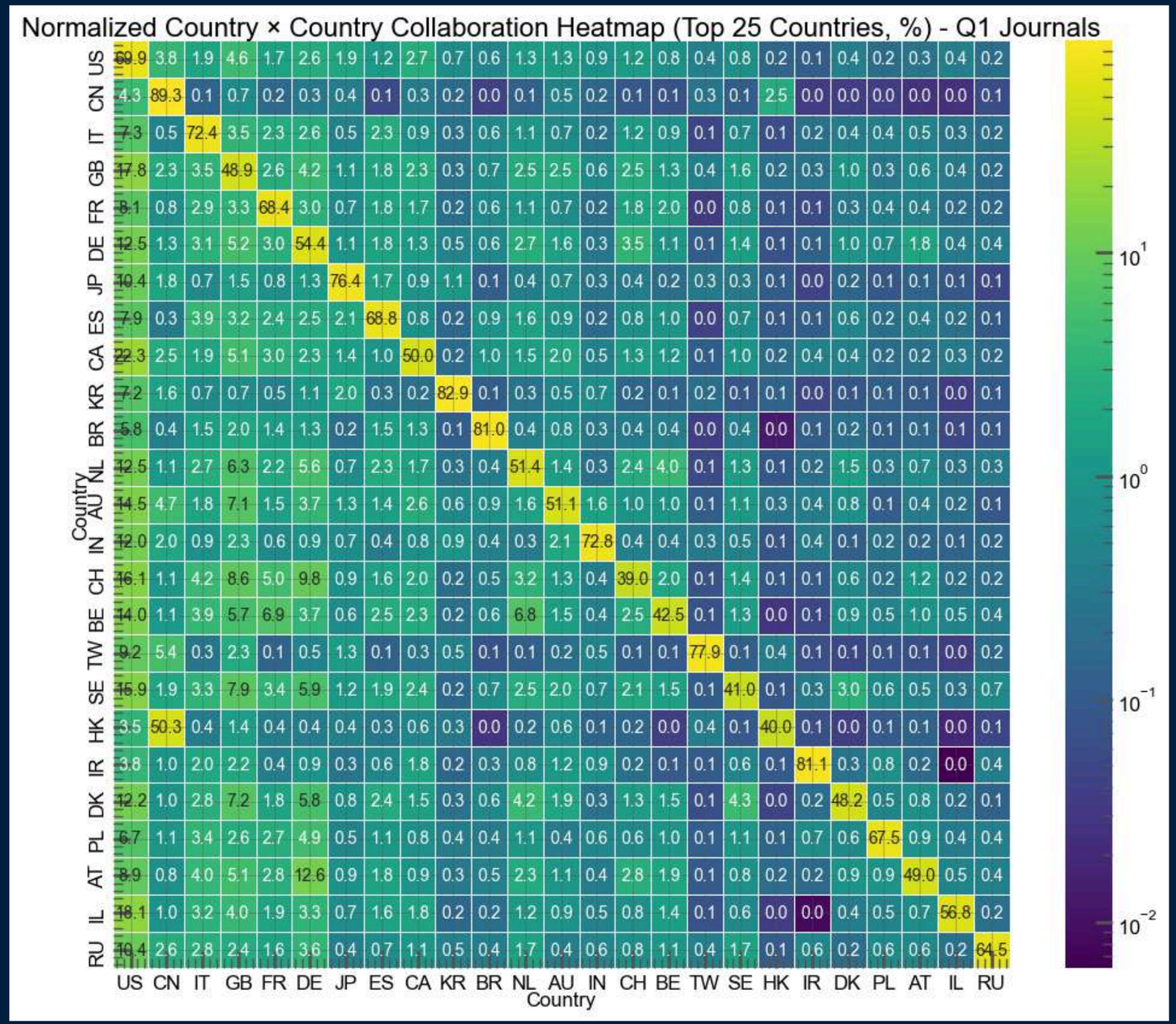


HSDSLab
Human & Social Data Science

Countries

Top Countries - Q1 Collaboration

- Rows = each country's collaboration profile → How this country distributes its collaborations across other countries. (Each number : % of the country's collabs)
- Columns = each country's importance as a partner → How much other countries collaborate with this country.

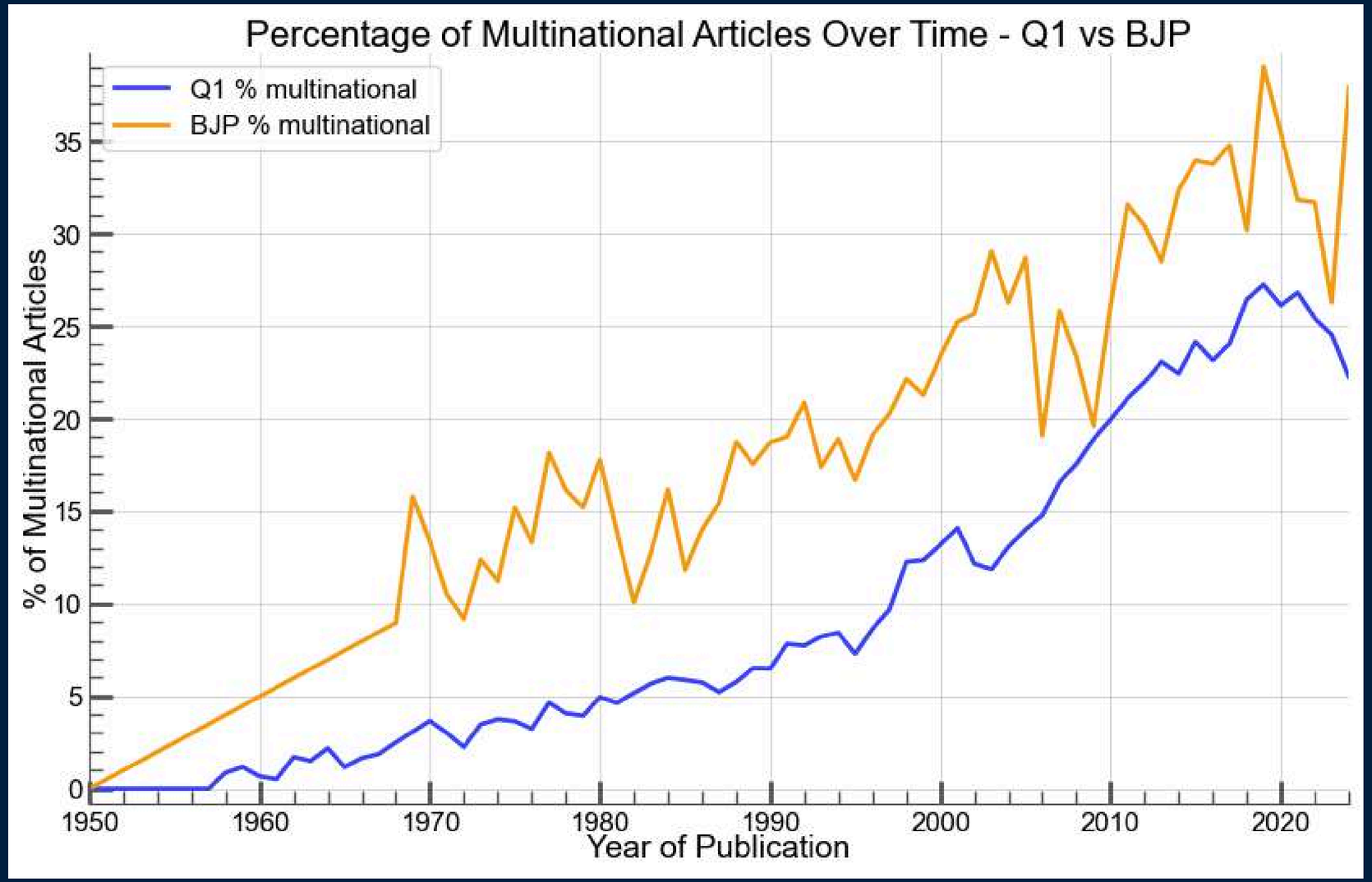


HSDSLab
Human & Social Data Science

Countries

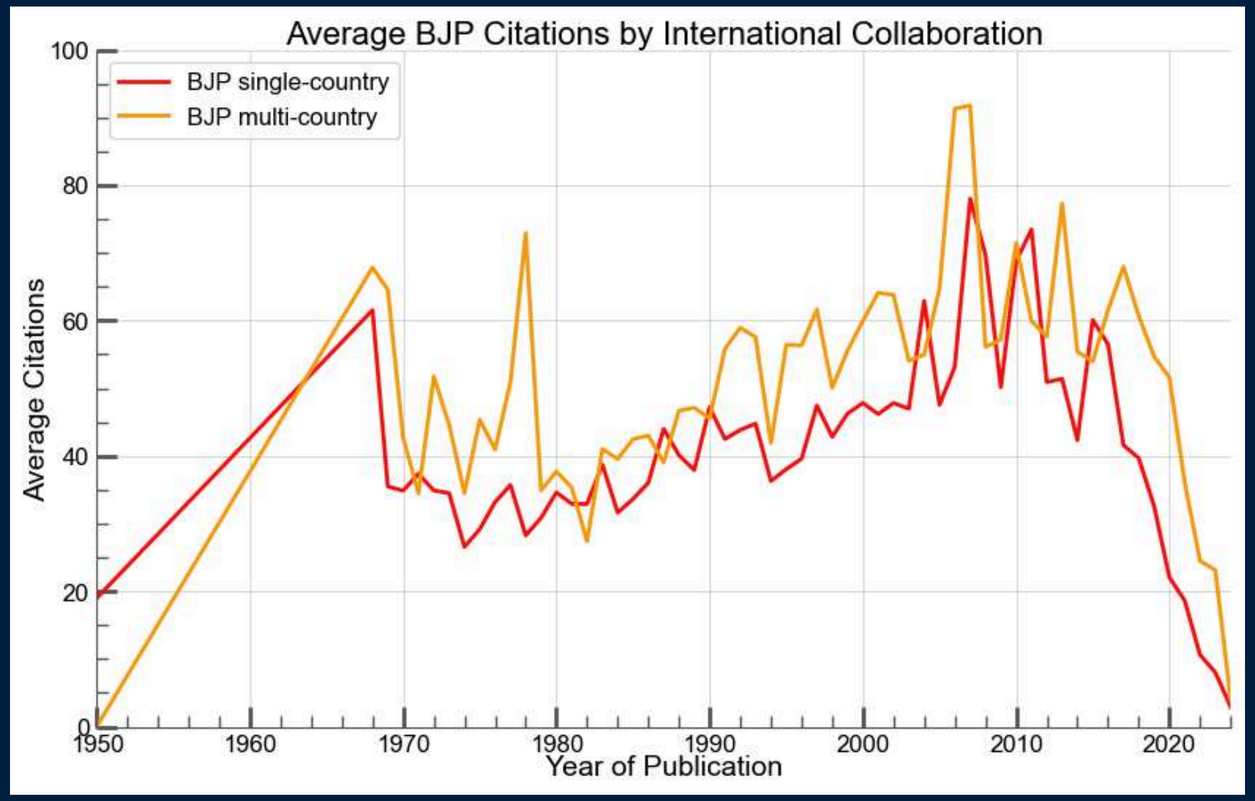
Multinational Articles

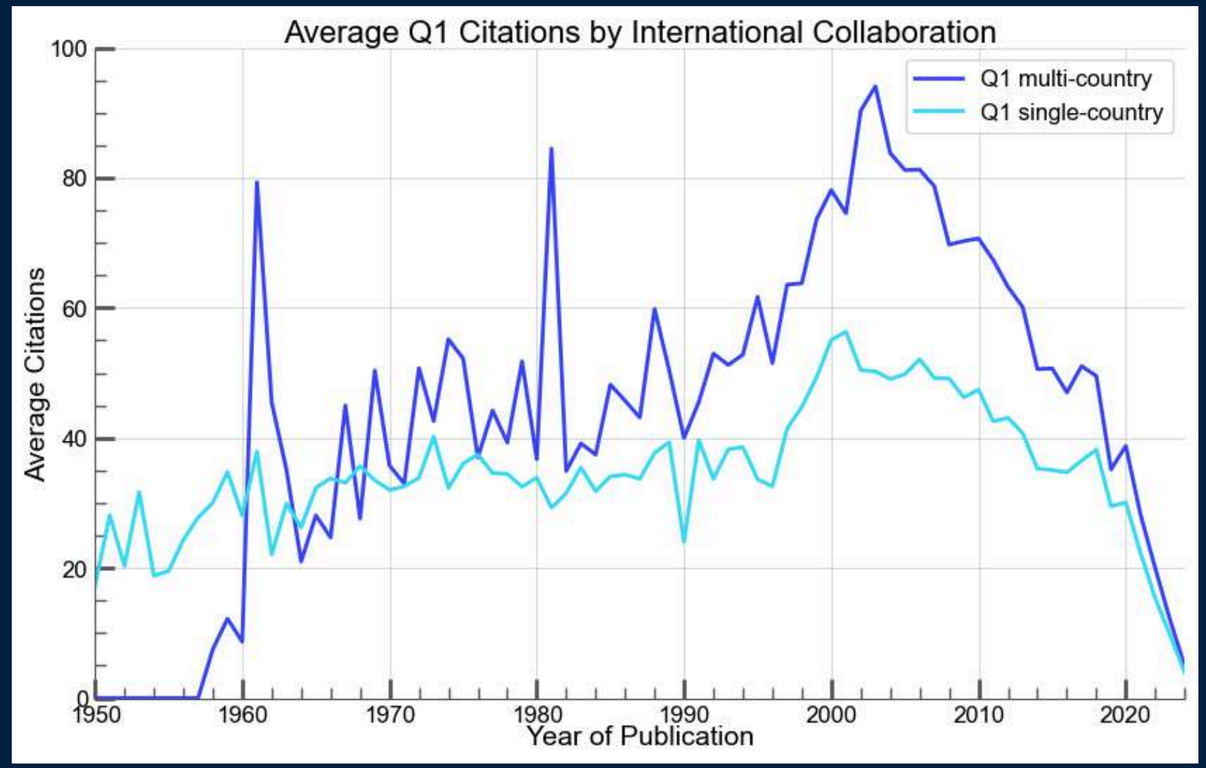
Works with authors from >= 2 DISTINCT countries



Countries

Multinational Articles - Related to citations





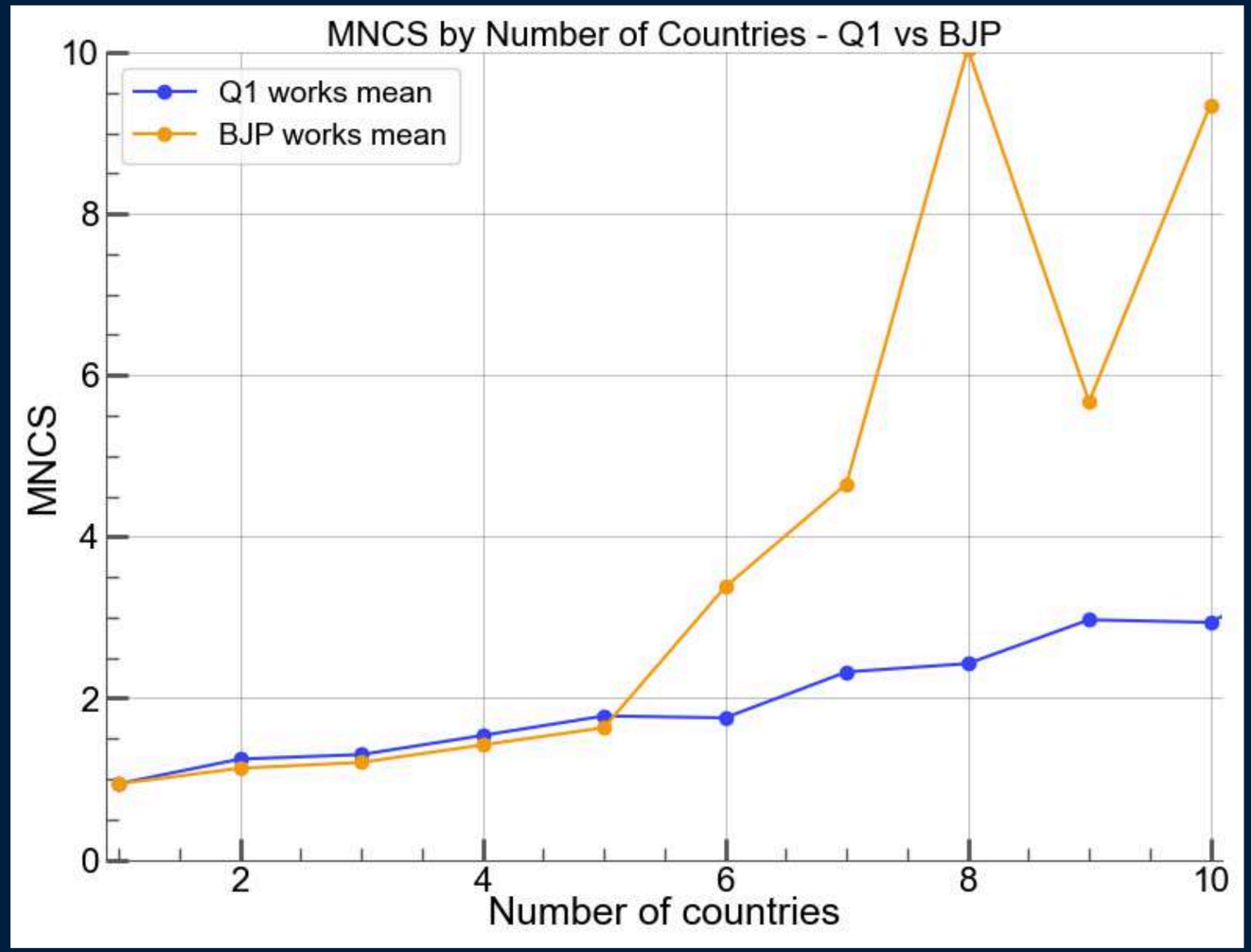
Countries

Related to MNCS

- Number of countries capped at 10 for clear visualization (too much variation otherwise)
- Small evolution, decent impact for this range
- Can be tested statistically (all numbers)

```
Pearson: r = 0.061, p-value = 0.000
Spearman: rho = 0.148, p-value = 0.000
```

→ Small but meaningful trend over a wider range

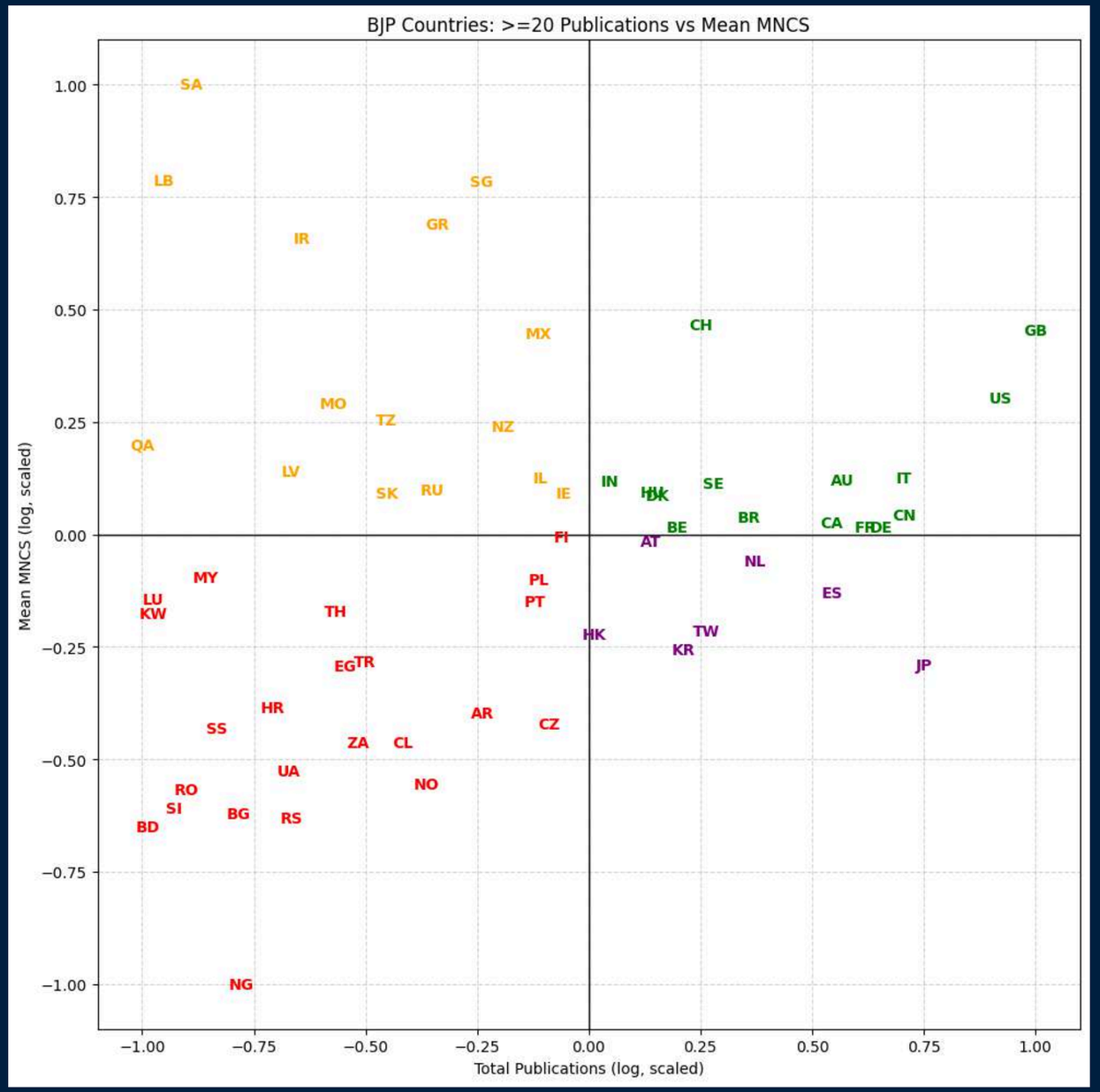


HSDSLab
Human & Social Data Science

Countries

Related to MNCS - BJP Countries

- High MNCS may depend on a few highly cited papers (SA, LB, SG)
- Major producers (US, GB) still show a high average MNCS
- Some major countries (ES, JP) show moderate MNCS.

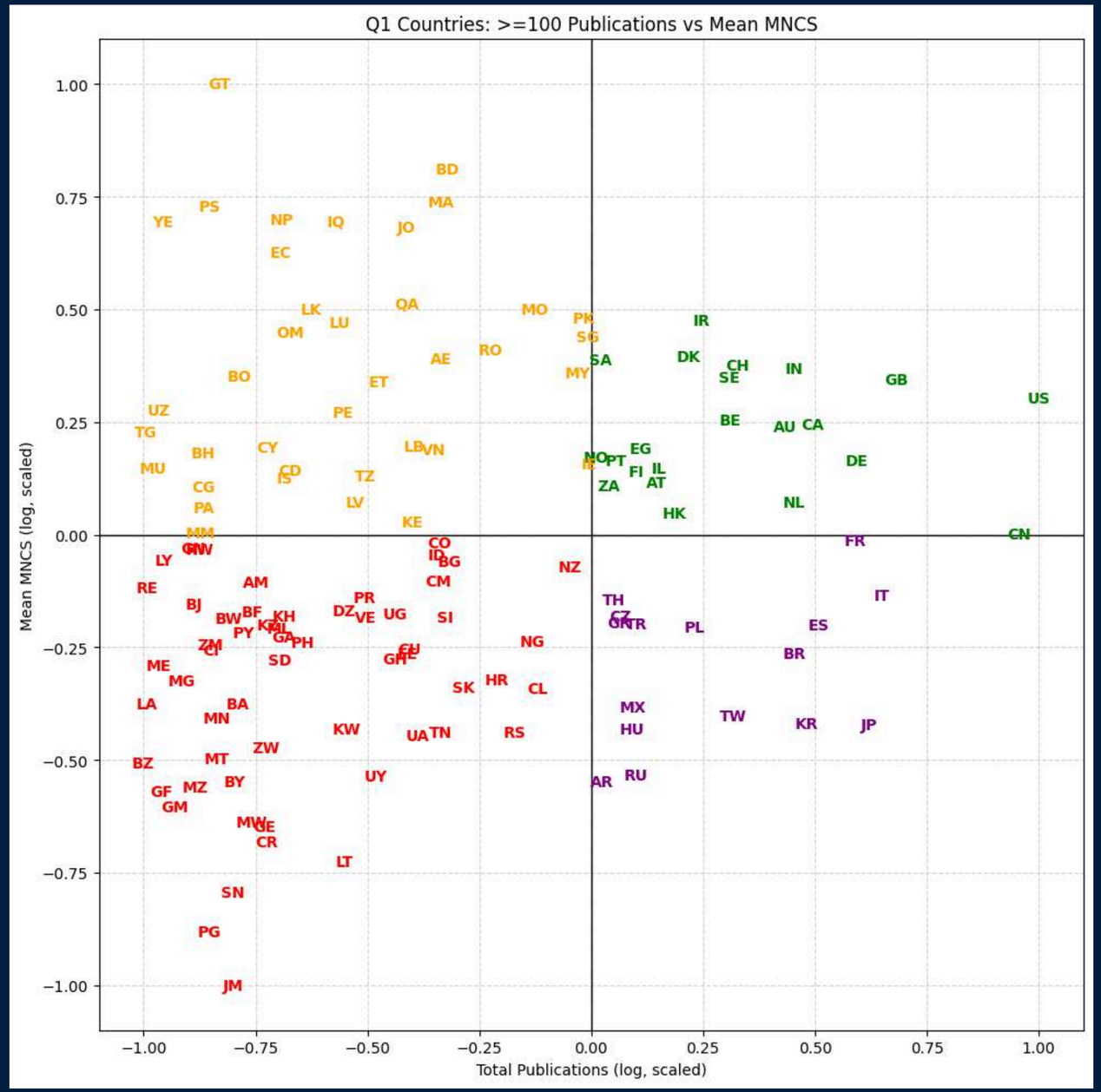


HSDSLab
Human & Social Data Science

Countries

Related to MNCS - Q1 Countries

- High MNCS may depend on a few highly cited papers (GT, BD, MA)
- Major producers (US, GB) still show a high average MNCS
- Some major countries (IT, JP) show moderate MNCS.



HSDSLab
Human & Social Data Science

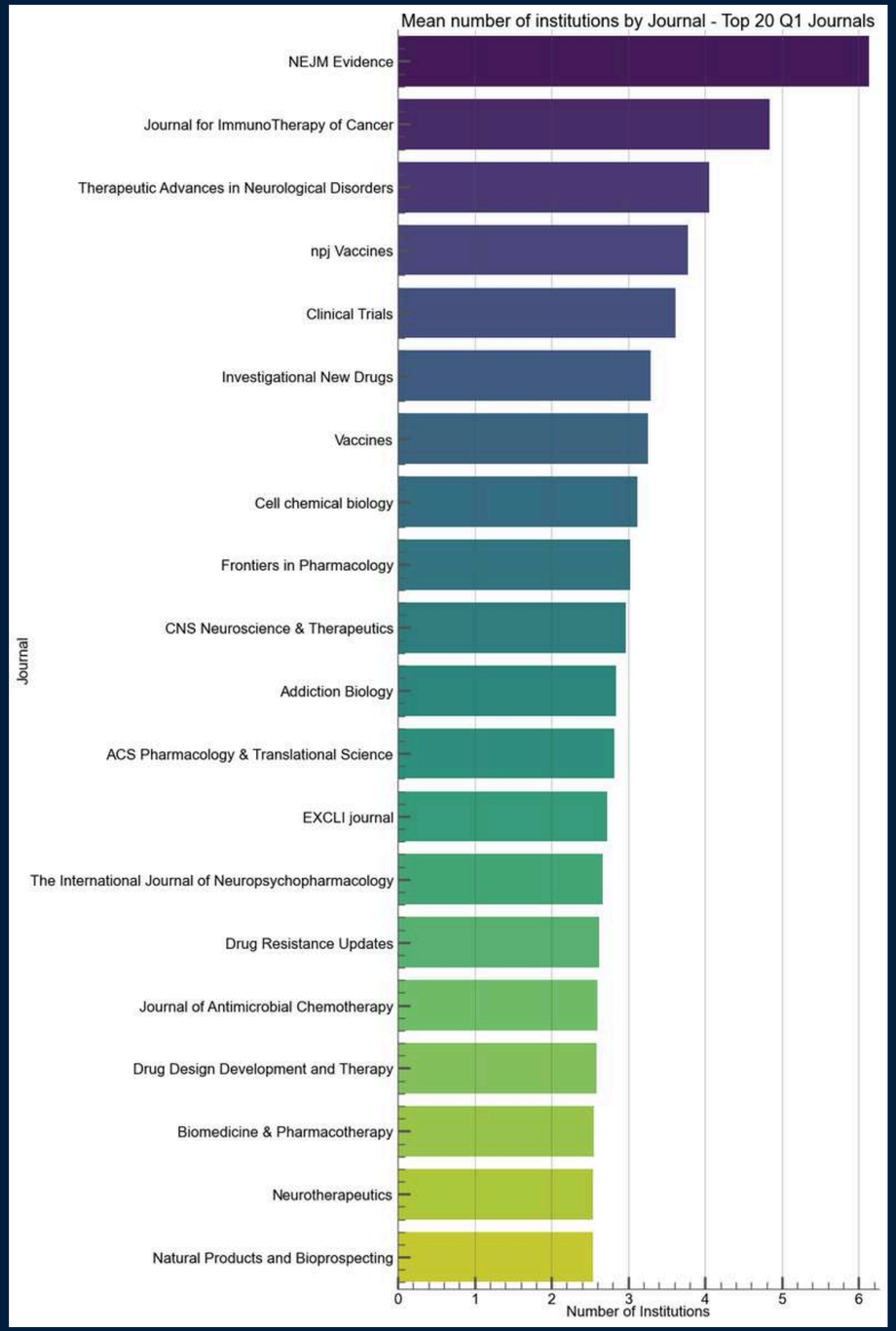
Institutions

Institutions

Journal Ranking

BJP is ranked **63rd**, with around

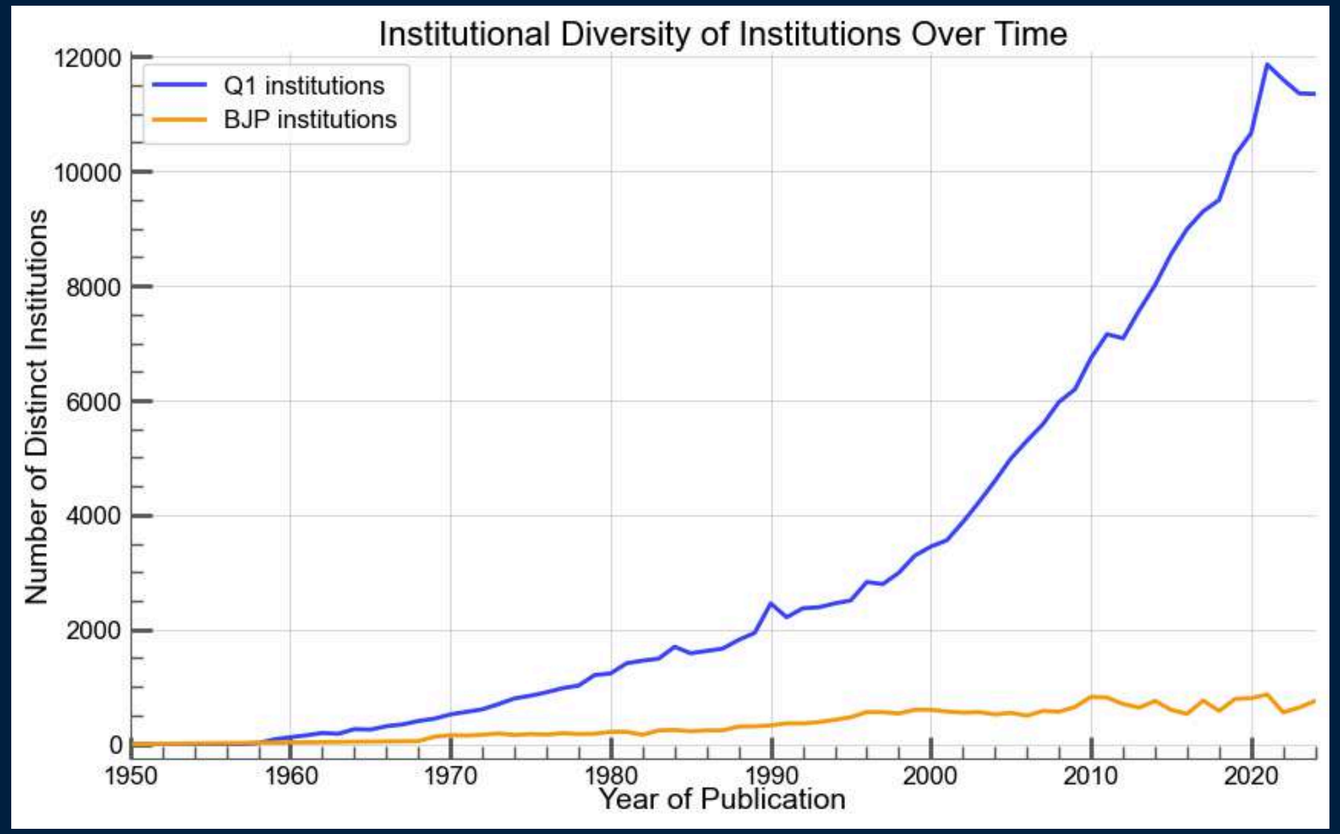
1.88 institutions per work.

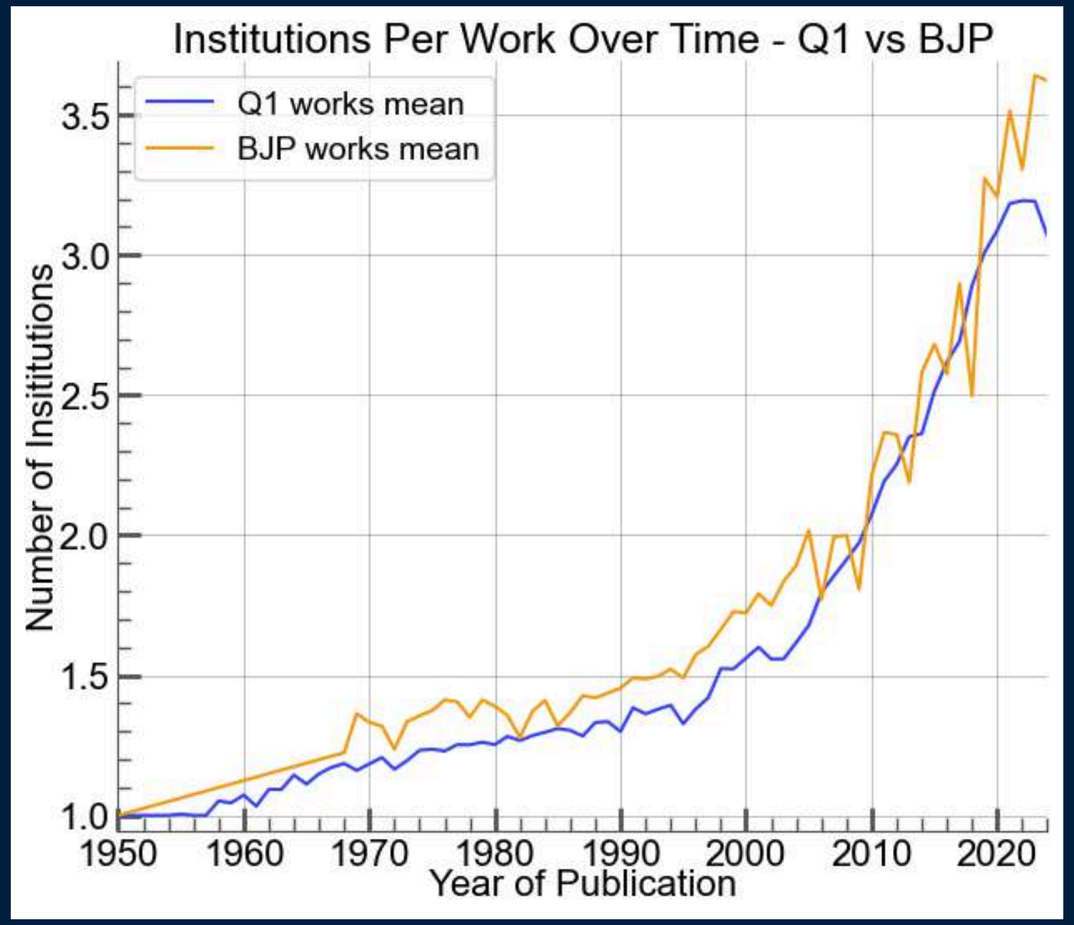


HSDSLab
Human & Social Data Science

Institutions

General Overview





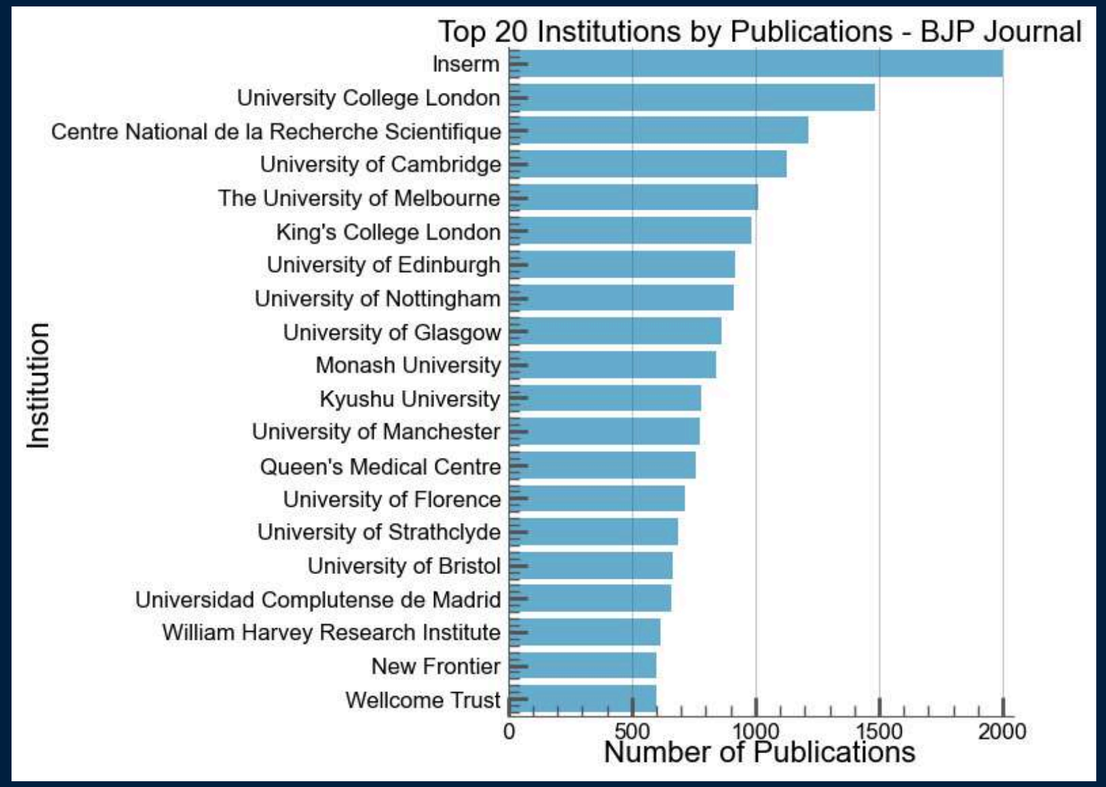
Peak in 2021 for BJP : 867 Institutions

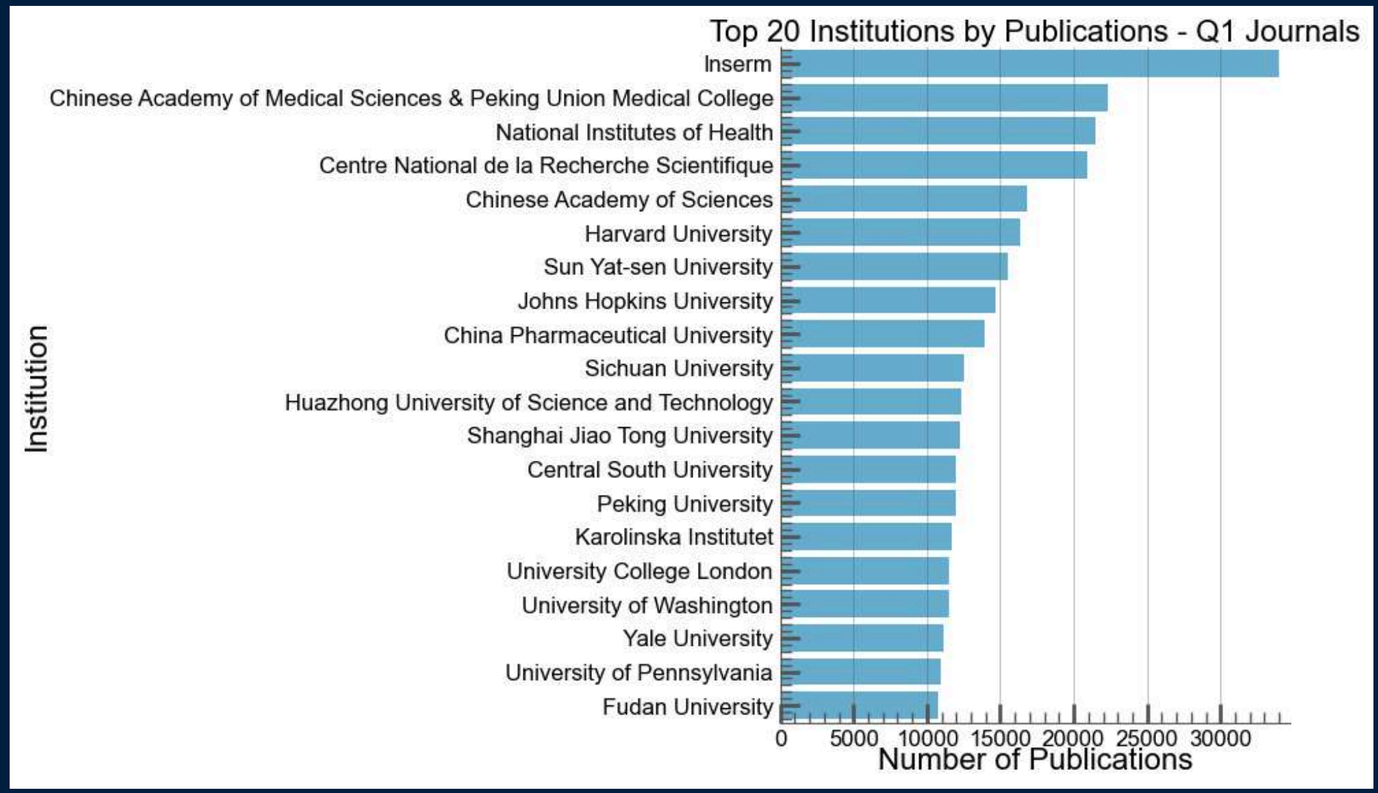
Peak in 2021 for Q1 Journals : 11864 Institutions

HSDSLab
Human & Social Data Science

Institutions

Top Institutions - Total Publications

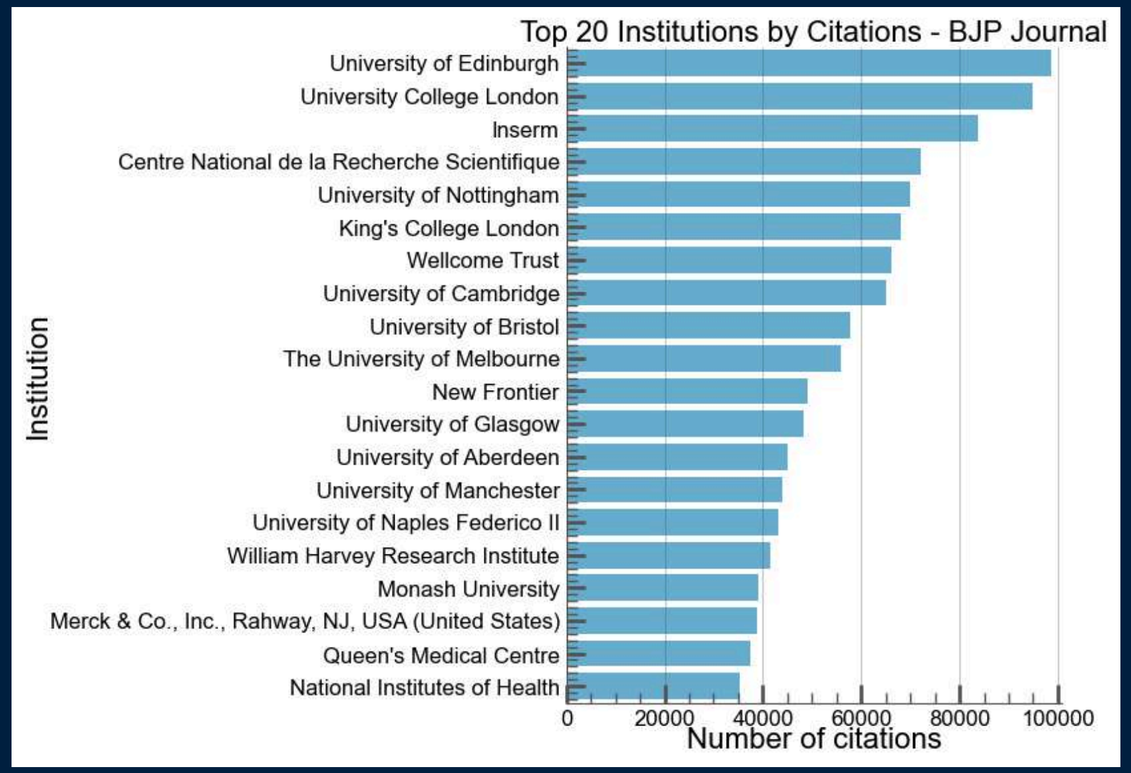


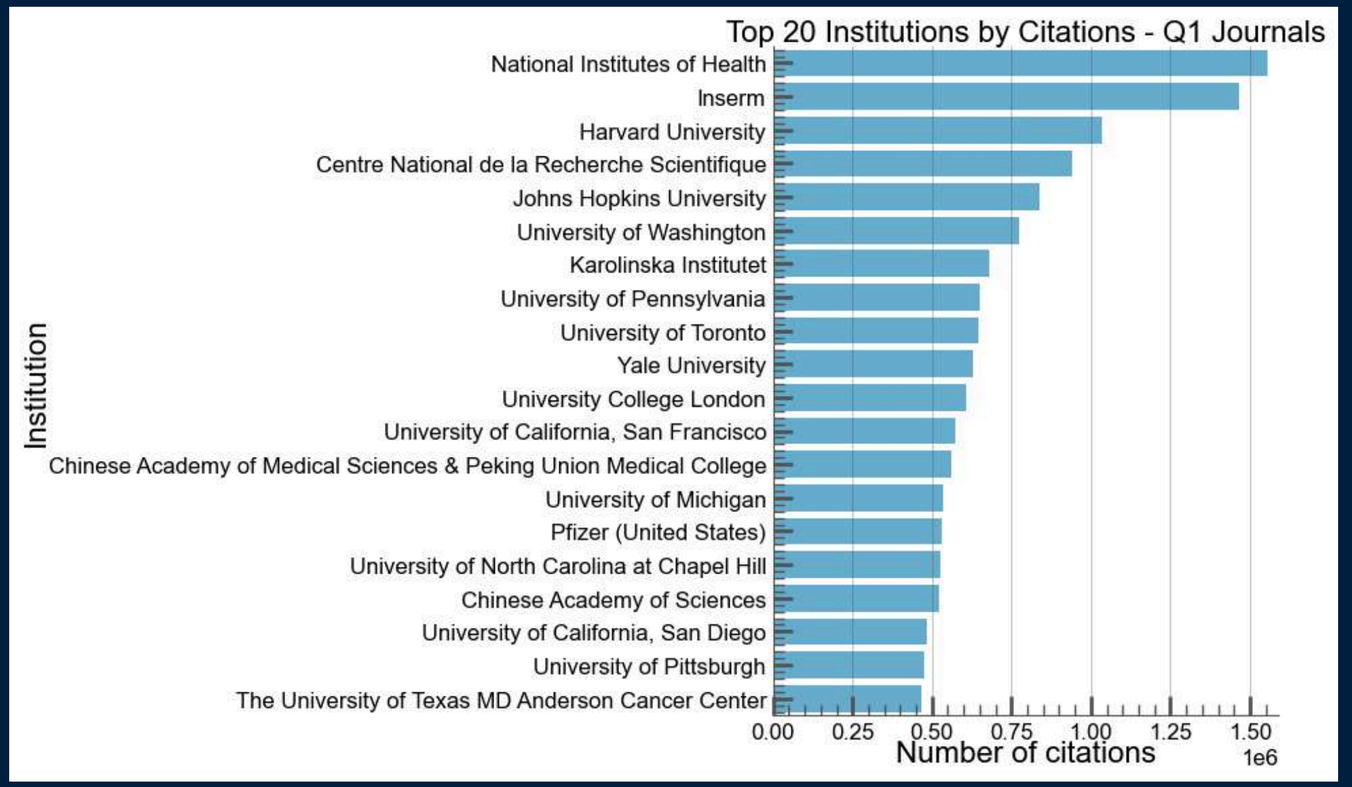


HSDSLab
Human & Social Data Science

Institutions

Top Institutions - Total Citations



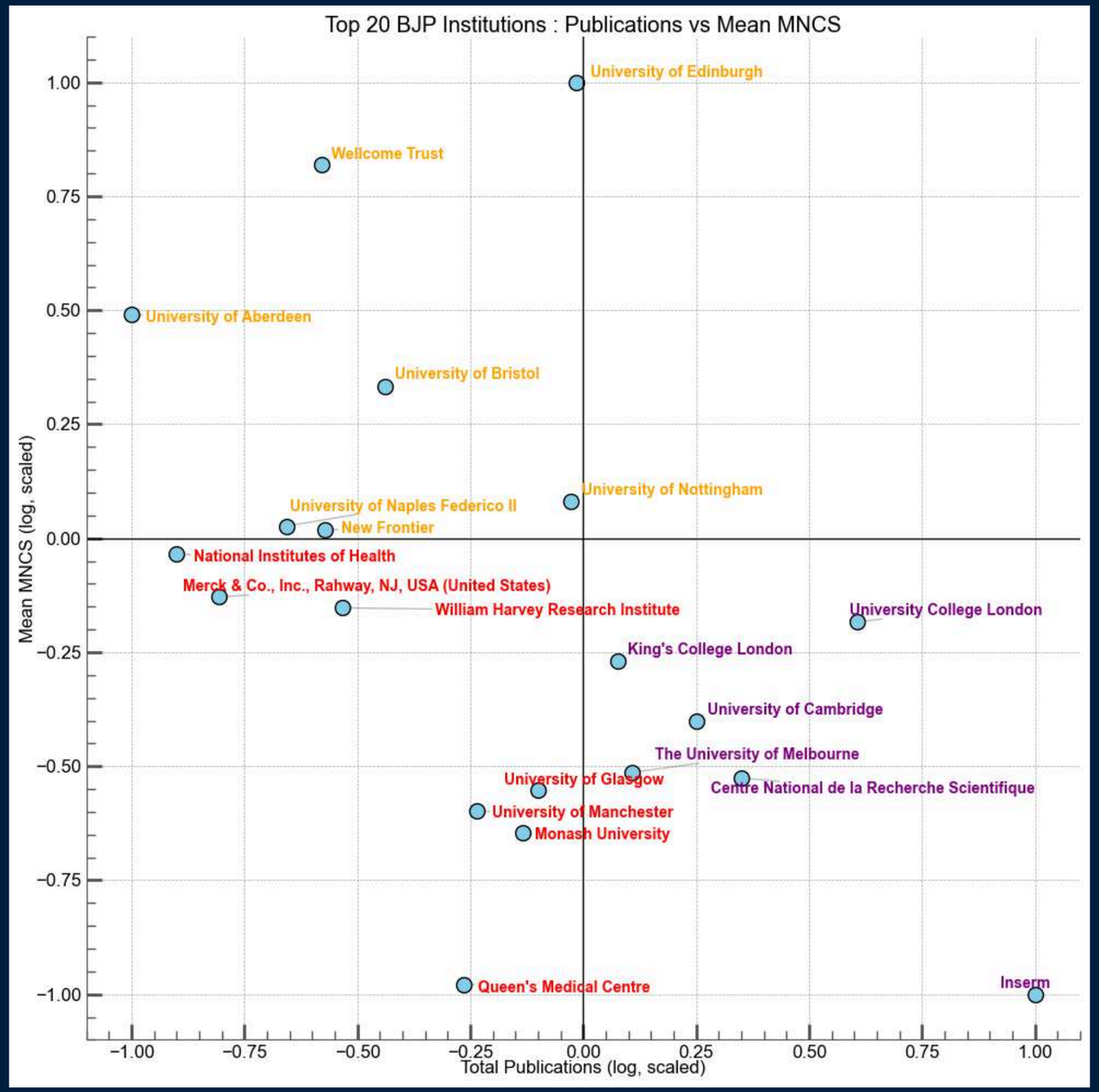


HSDSLab
Human & Social Data Science

Institutions

Related to MNCS - BJP Institutions

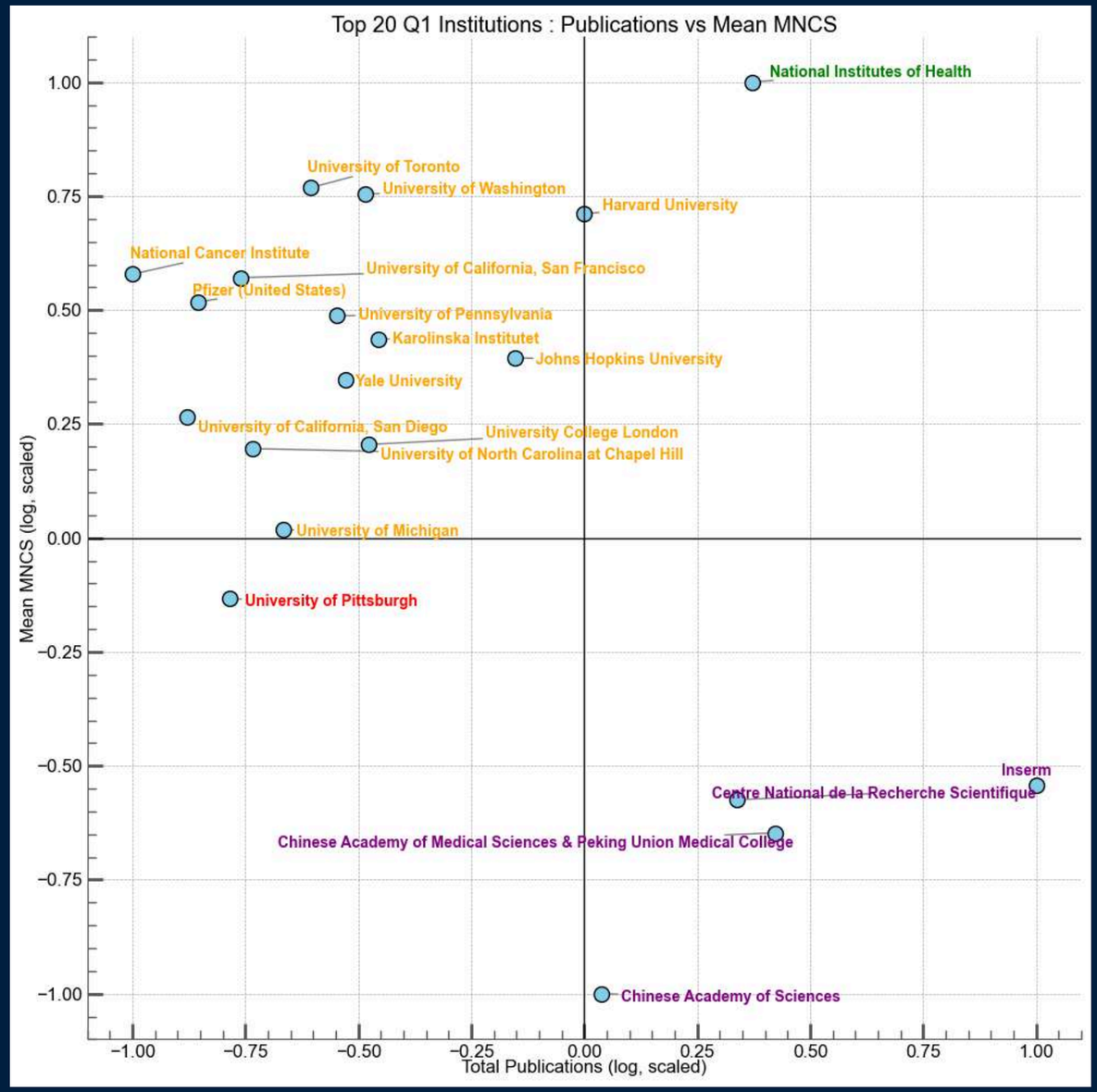
- Institutions with more publications tend to have a lower MNCS compared to other top institutions
- High MNCS despite many publications (Univ. of Edinburgh, Univ. College London)
- Low MNCS despite fewer publications (Queen's Medical Centre, Monash University)



Institutions

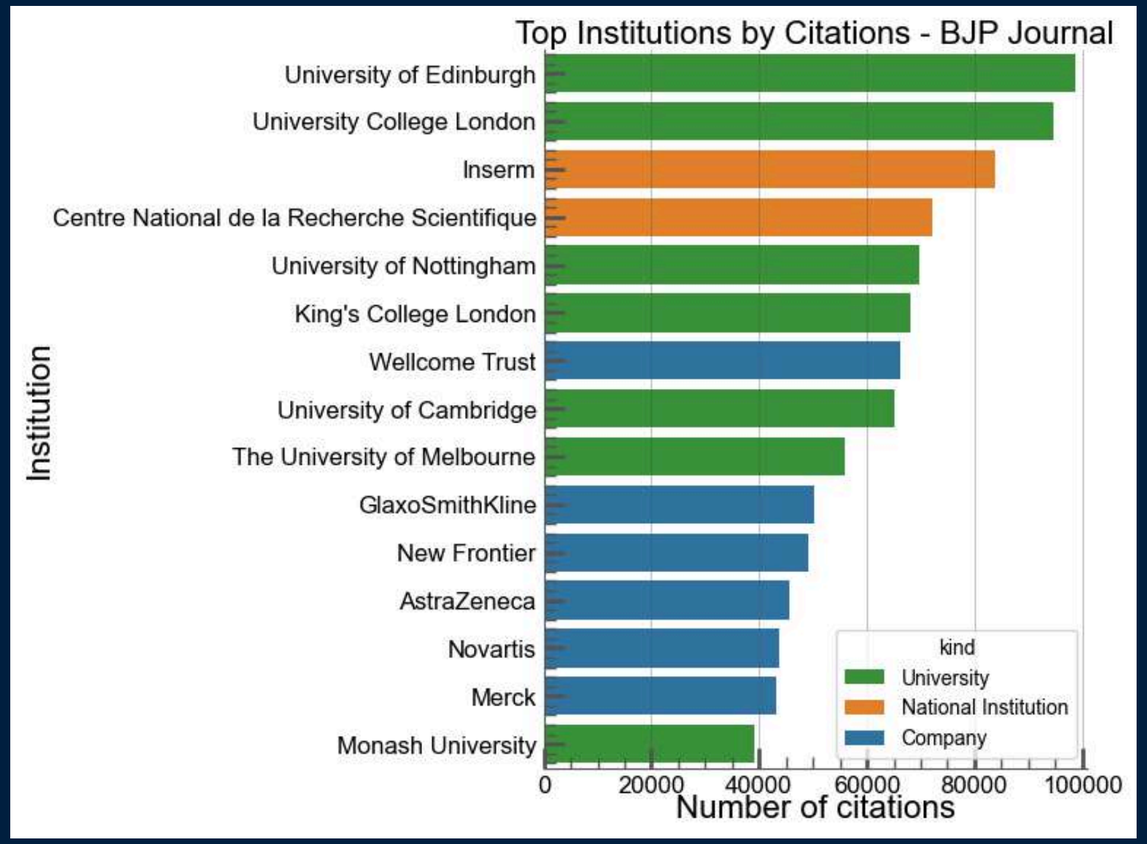
Related to MNCS - Q1 Institutions

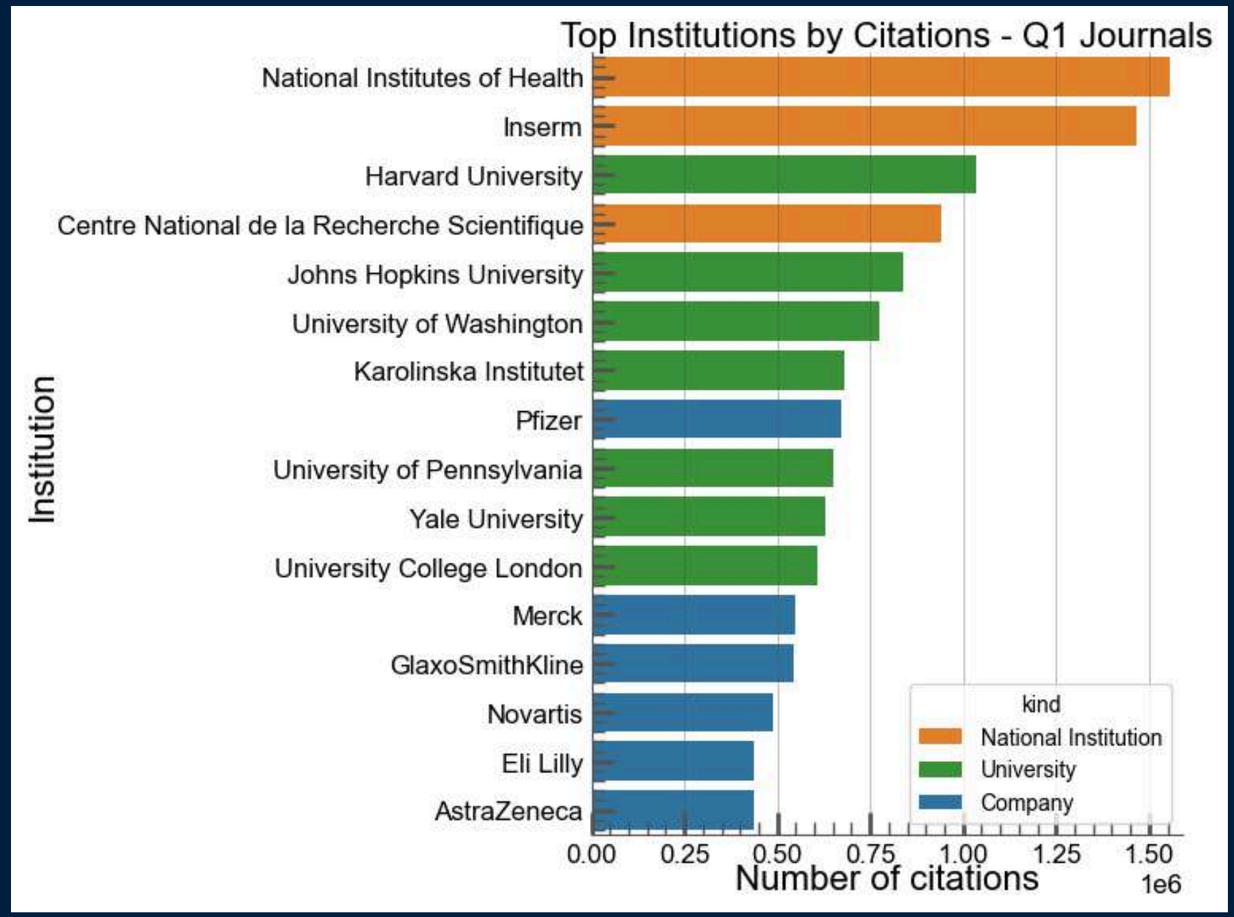
- Institutions with more publications tend to have a lower MNCS compared to other top institutions
- High MNCS despite many publications (National Instit. of Health, Harvard University)
- Low MNCS despite fewer publications (Chinese Academy of Sciences, Univ. of Pittsburgh)



Institutions

Top Institutions - Total Citations (Type)

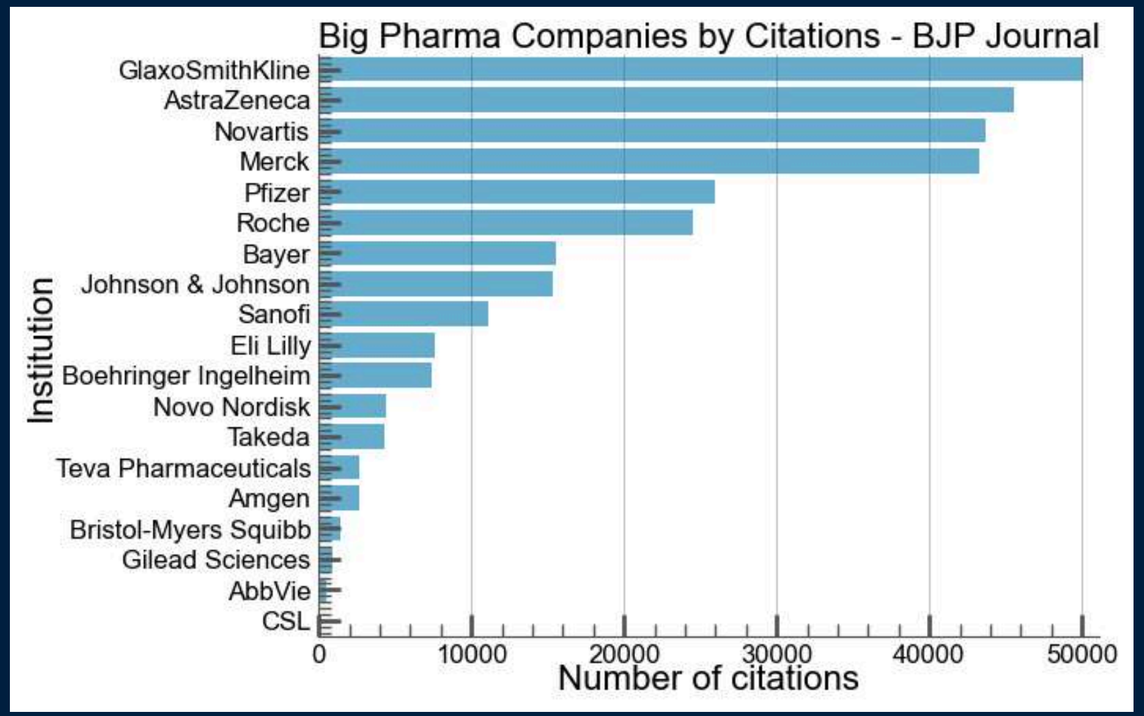


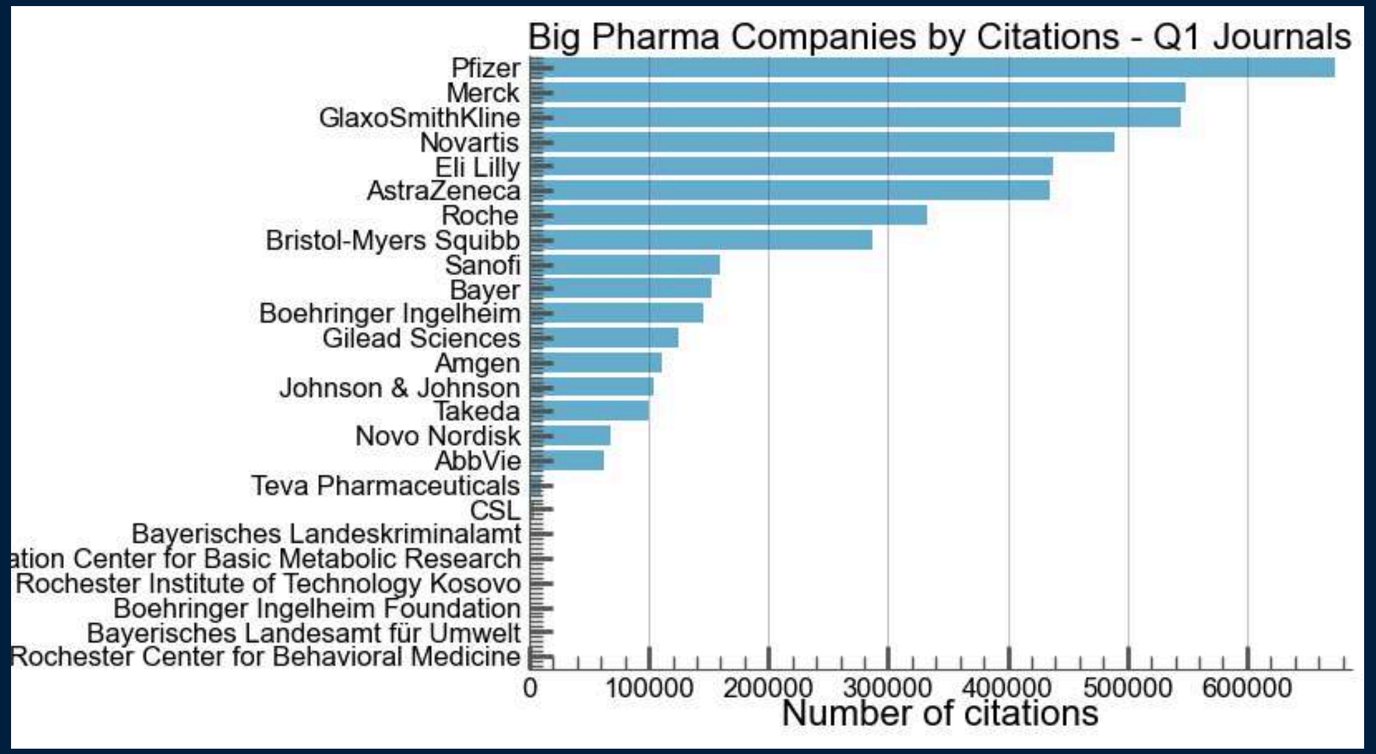


HSDSLab
Human & Social Data Science

Institutions

Top Institutions - Total Citations (Big Pharma Companies)

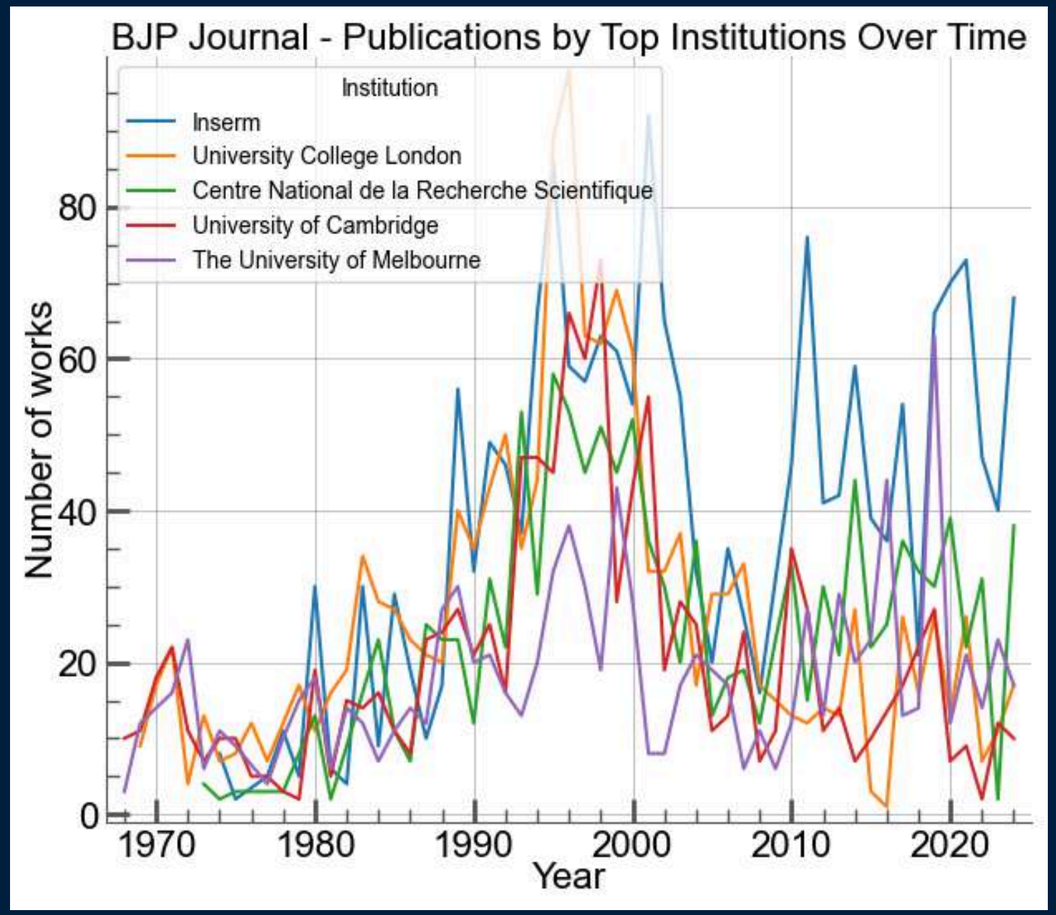


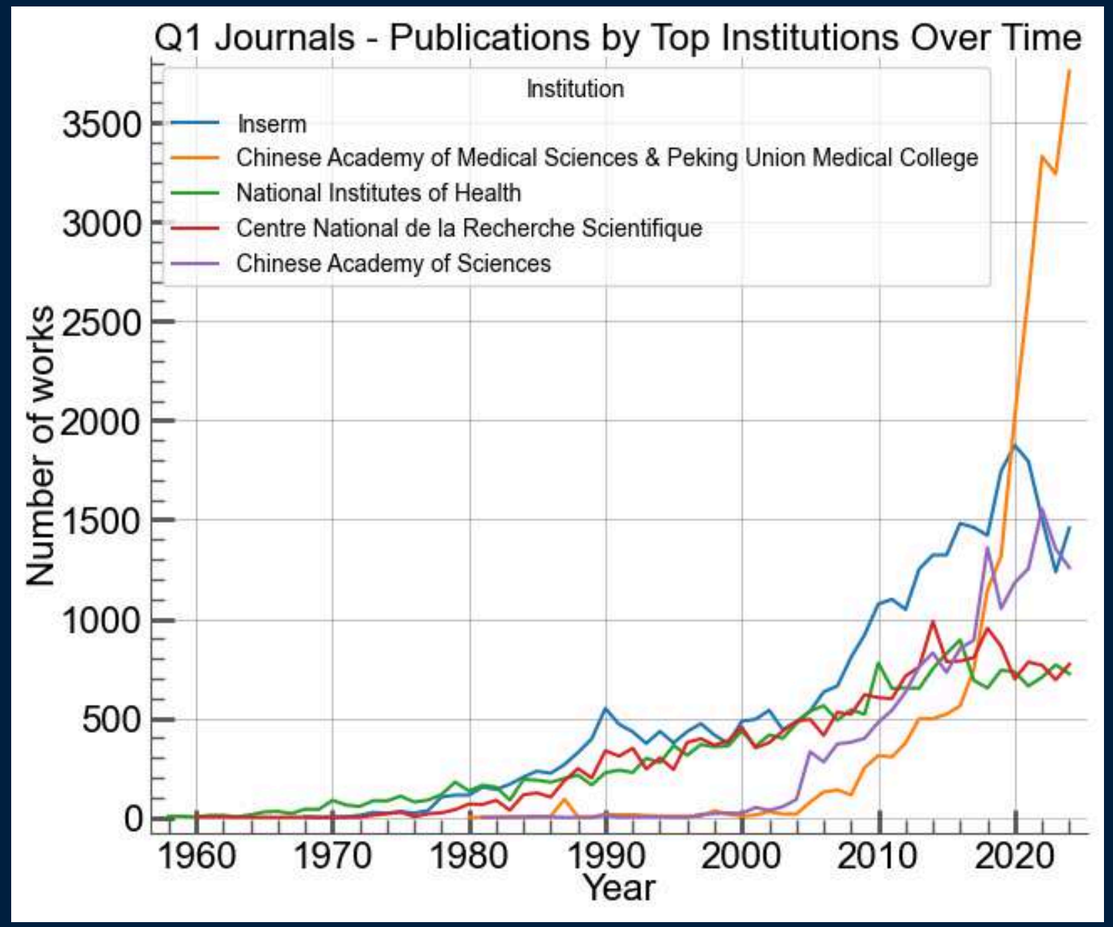


HSDSLab
Human & Social Data Science

Institutions

Top Institutions - Evolution of Publications



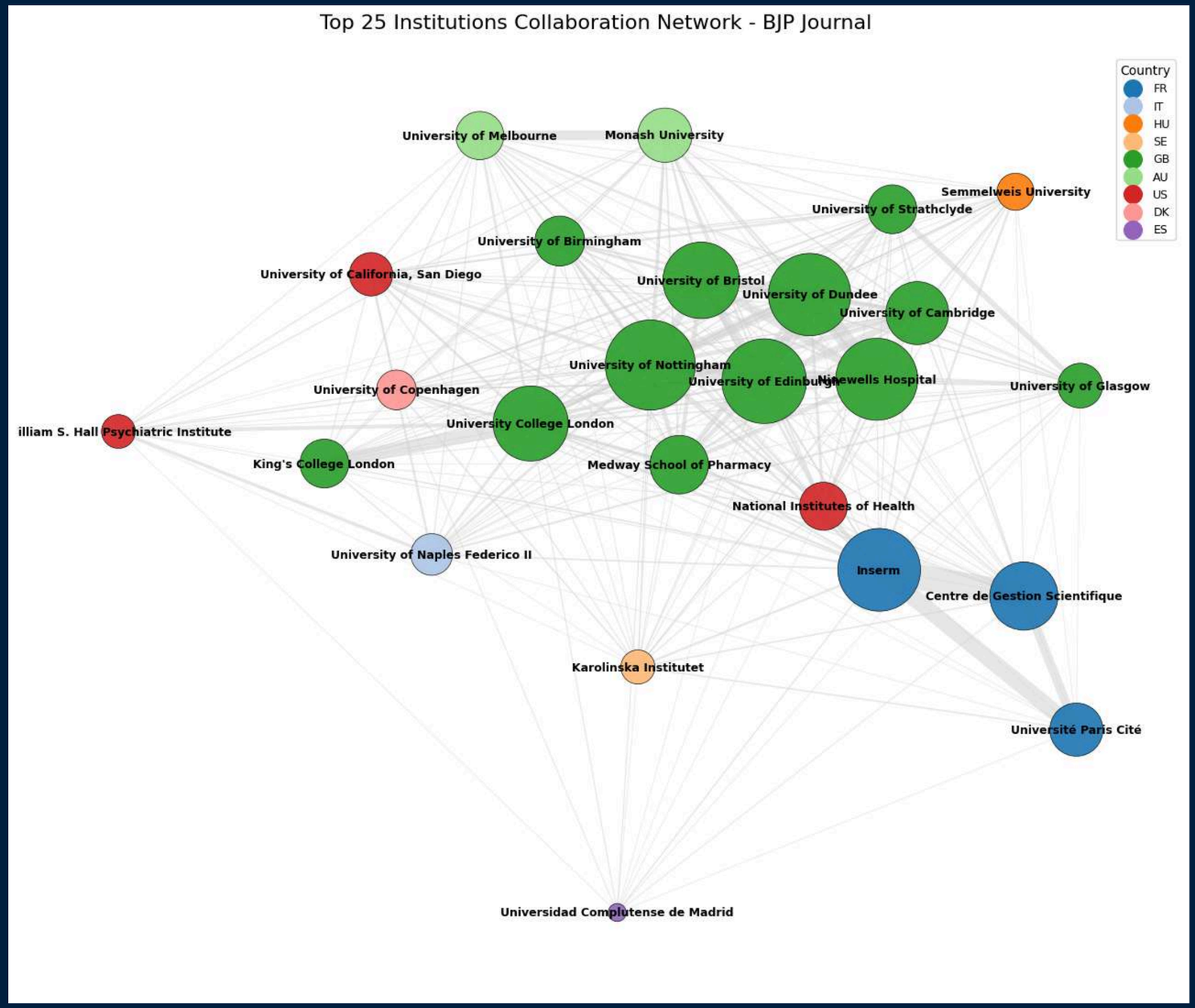


HSDSLab
Human & Social Data Science

Institutions

Top Institutions - Collaboration Network - BJP

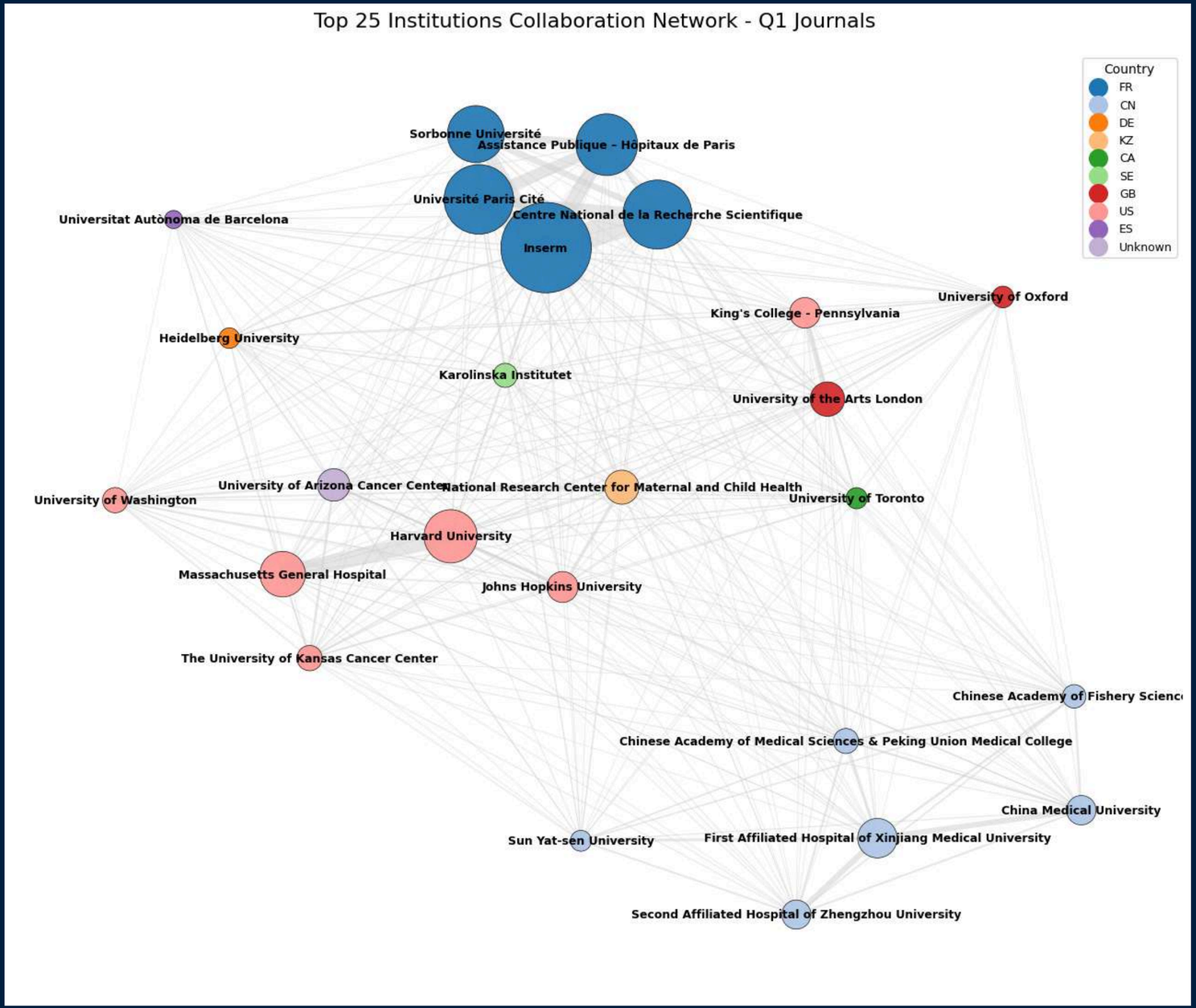
- Selected the top 25 institutions by total number of collaborations.
- Collaboration = two institutions co-authoring the same article; counted once per article.
- Institutions from same country seem closer, (only european) high influence from british institutions



Institutions

Top Institutions - Collaboration Network - Q1

- Selected the top 25 institutions by total number of collaborations.
- Collaboration = two institutions co-authoring the same article; counted once per article.
- Institutions from same country, continent seem closer, high influence from french institutions



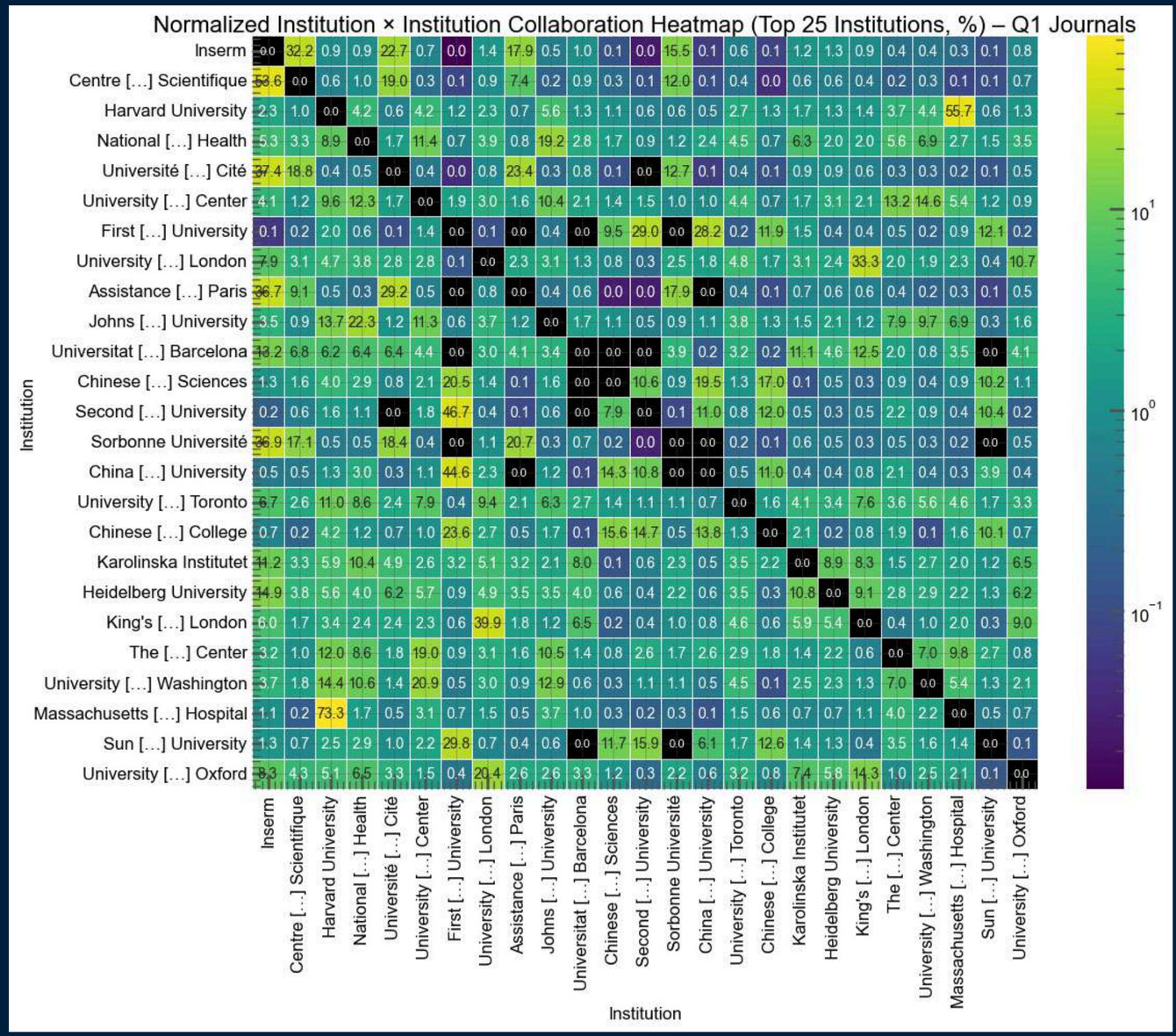
HSDSLab
Human & Social Data Science

Institutions

Top Institutions - BJP Collaboration

- Rows = each institution's collaboration profile → How this country distributes its collaborations across other countries. (Each number : % of the institution's collabs)

- Columns = each institution's importance as a partner → How much other countries collaborate with this country.



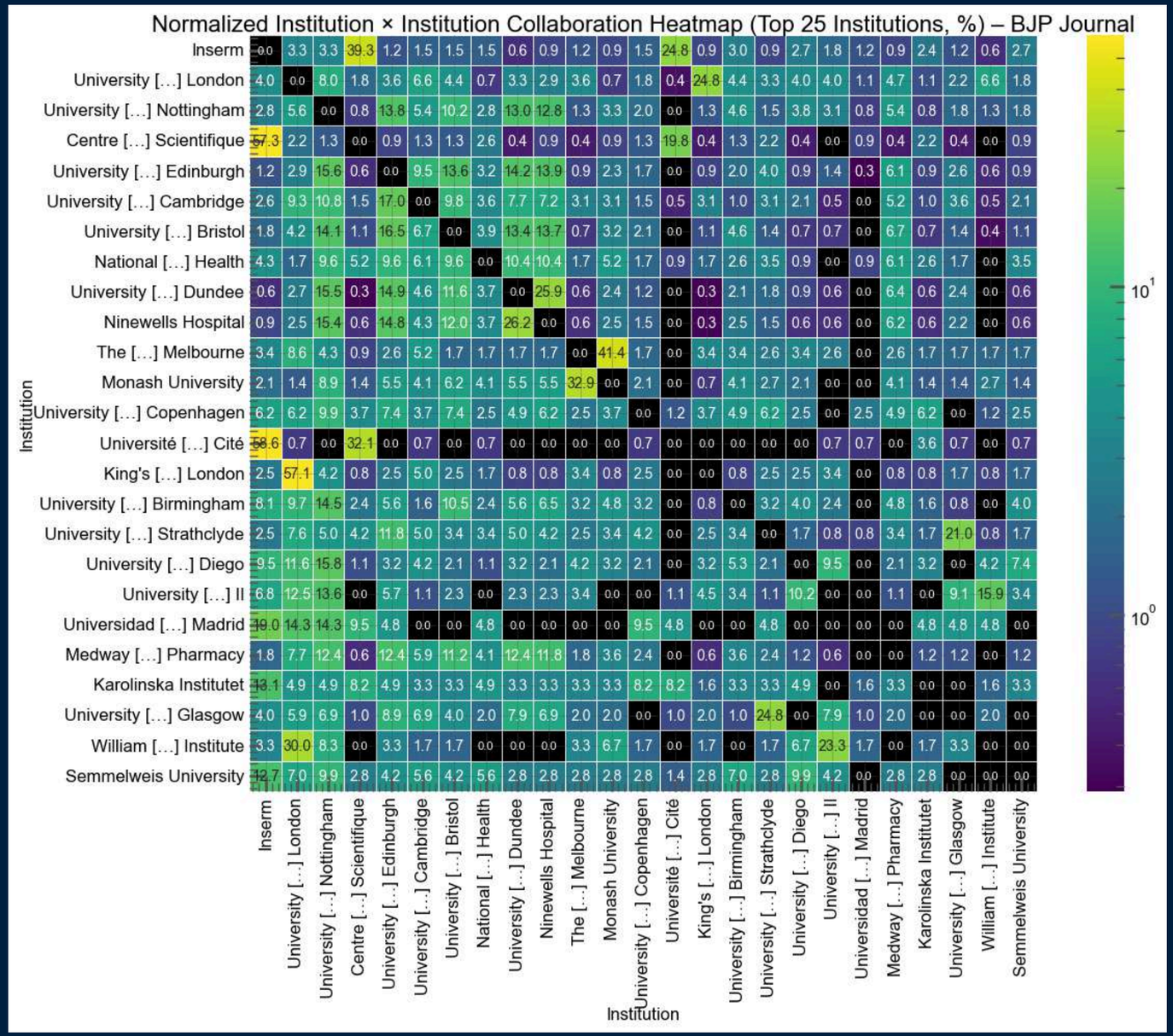
HSDSLab
Human & Social Data Science

Institutions

Top Institutions - Q1 Collaboration

- Rows = each institution's collaboration profile → How this institution distributes its collaborations across other institutions (Each number : % of the institution's collabs)

- Columns = each institution's importance as a partner → How much other institutions collaborate with this institution

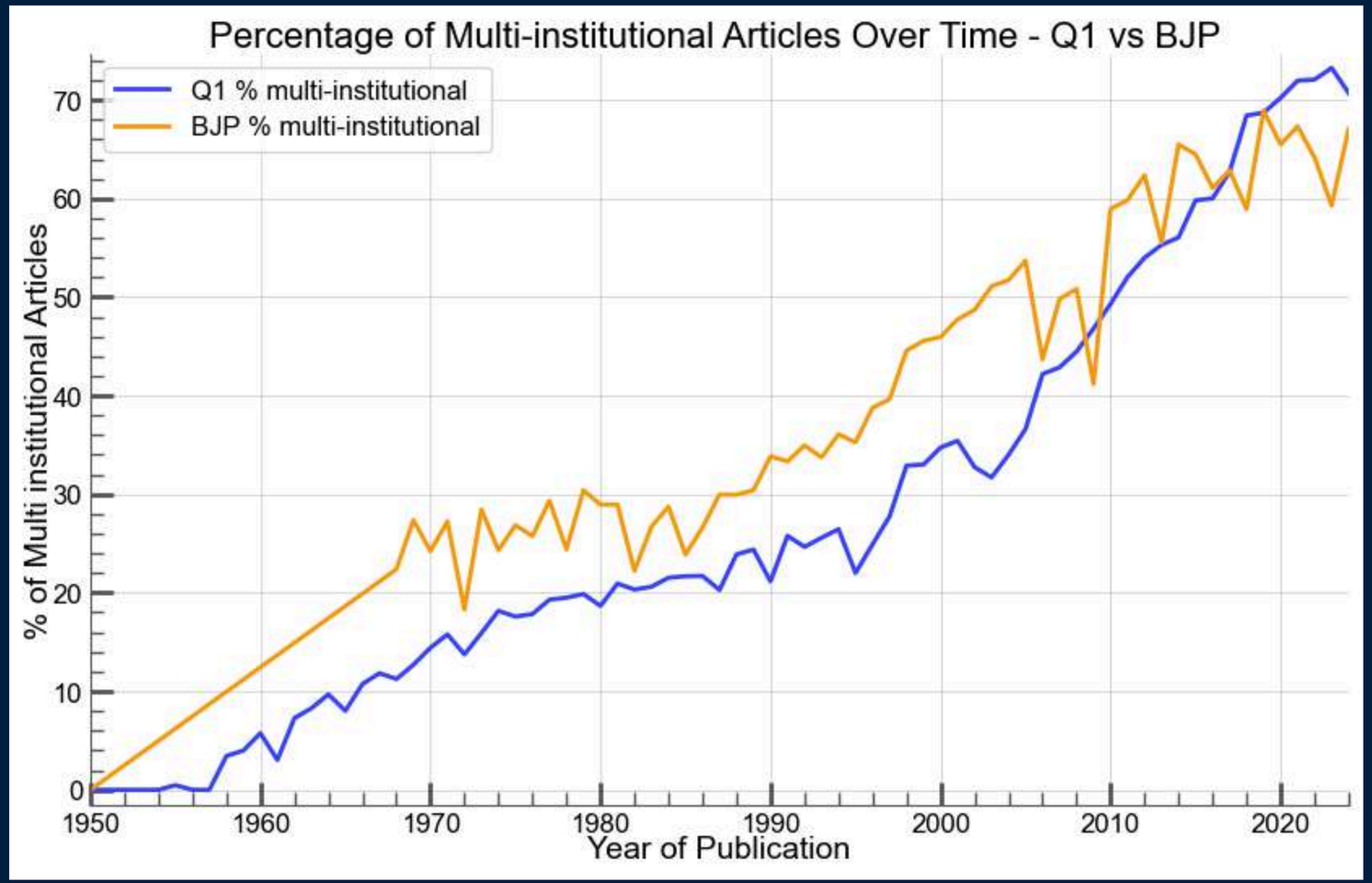


HSDSLab
Human & Social Data Science

Institutions

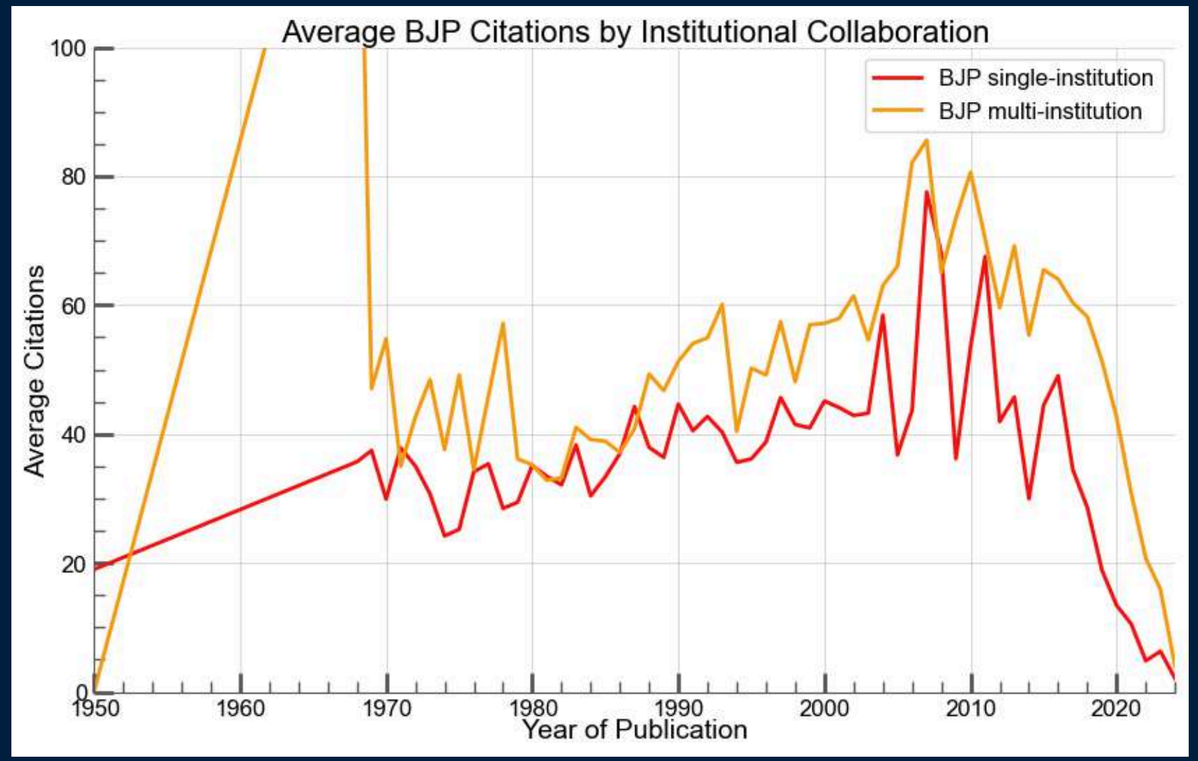
Multi-Institutional Articles

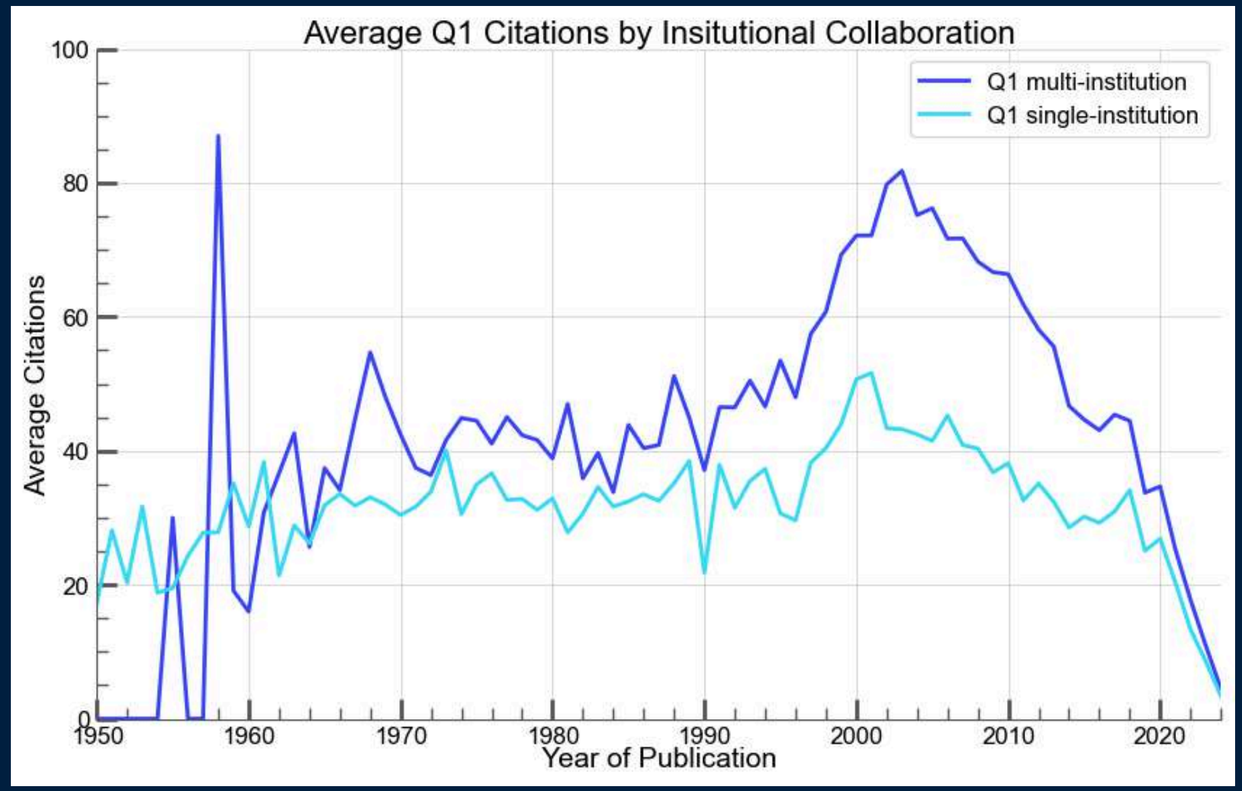
Works with >= 2 DISTINCT Institutions



Institutions

Multi-Institutional Articles - Related to citations





HSDSLab
Human & Social Data Science

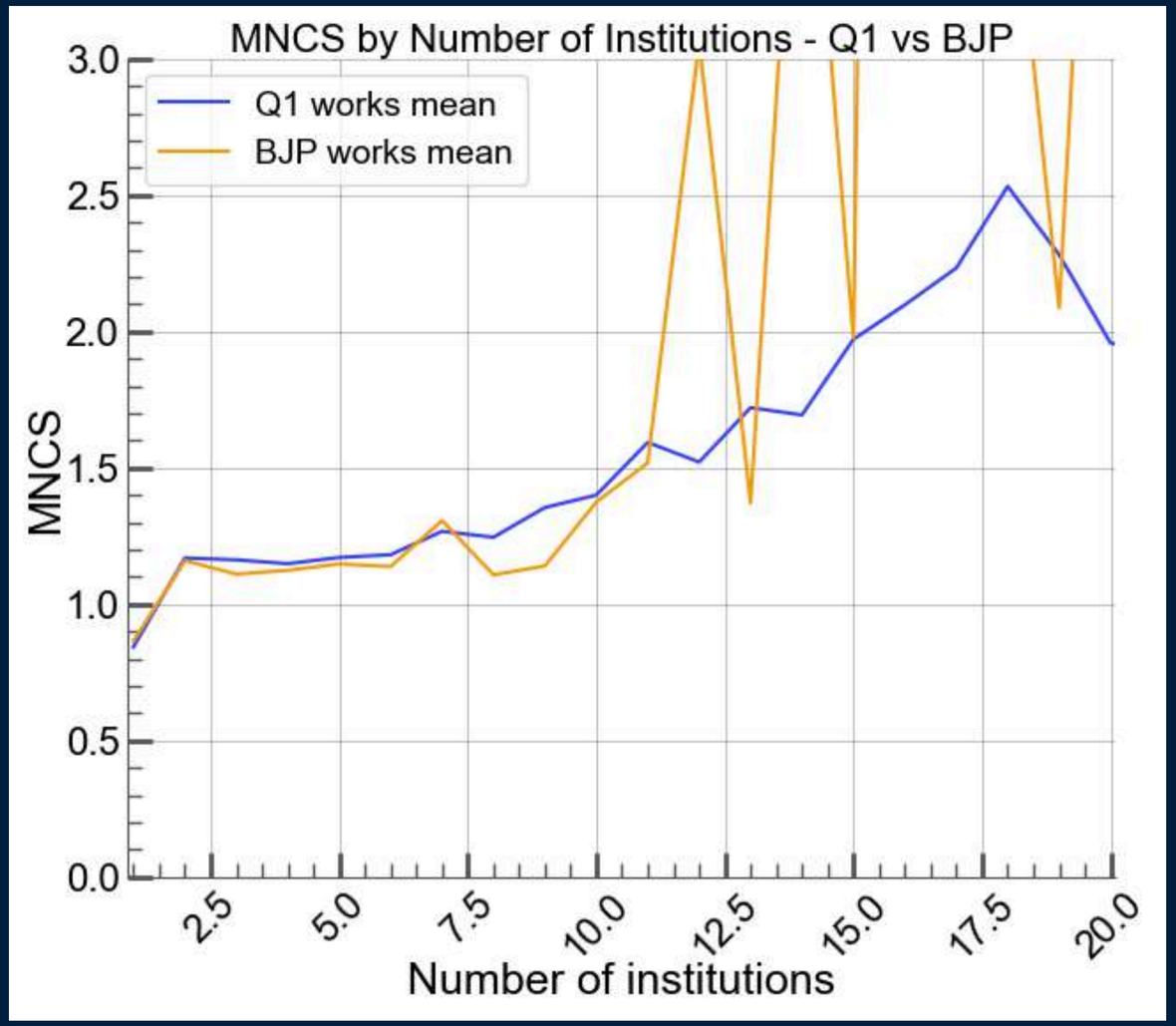
Institutions

Related to MNCS

- Number of institutions capped at 20 for clear visualization (too much variation otherwise)
- Small evolution since >=8 Institutions, decent impact
- Can be tested statistically (all numbers)

```
Pearson: r = 0.061, p-value = 0.000
Spearman: rho = 0.252, p-value = 0.000
```

→ Small but meaningful trend over a wider range



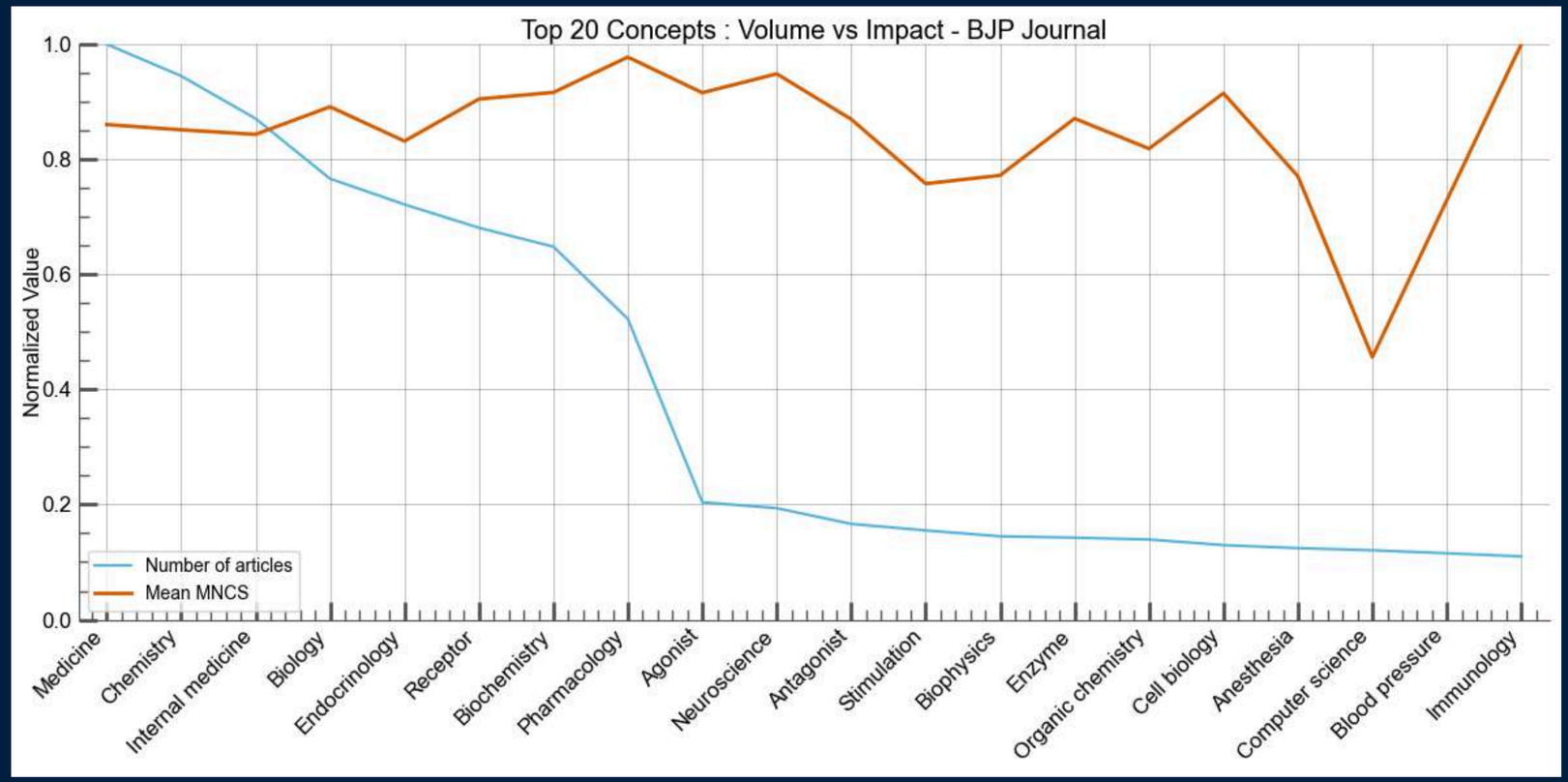
HSDSLab
Human & Social Data Science

Topics

Topics

Top Concepts - BJP Publications vs MNCS

- Concepts are variables from OpenAlex.
- They represent scientific topics or subjects associated with publications, authors, or journals.
- Each article can be linked to multiple concepts

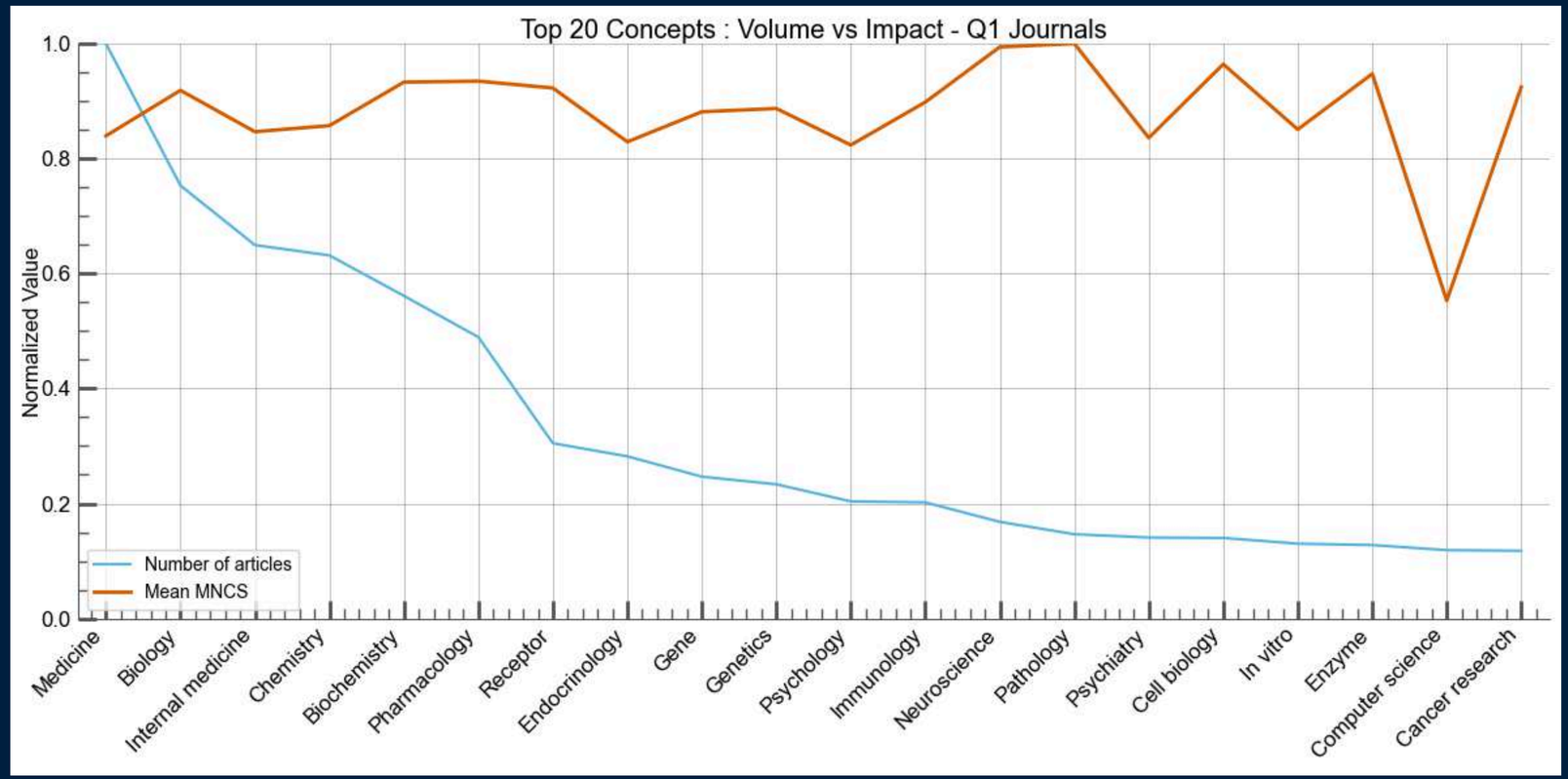


HSDSLab
Human & Social Data Science

Topics

Top Concepts - Q1 Publications vs MNCS

- Concepts are variables from OpenAlex.
- They represent scientific topics or subjects associated with publications, authors, or journals.
- Each article can be linked to multiple concepts

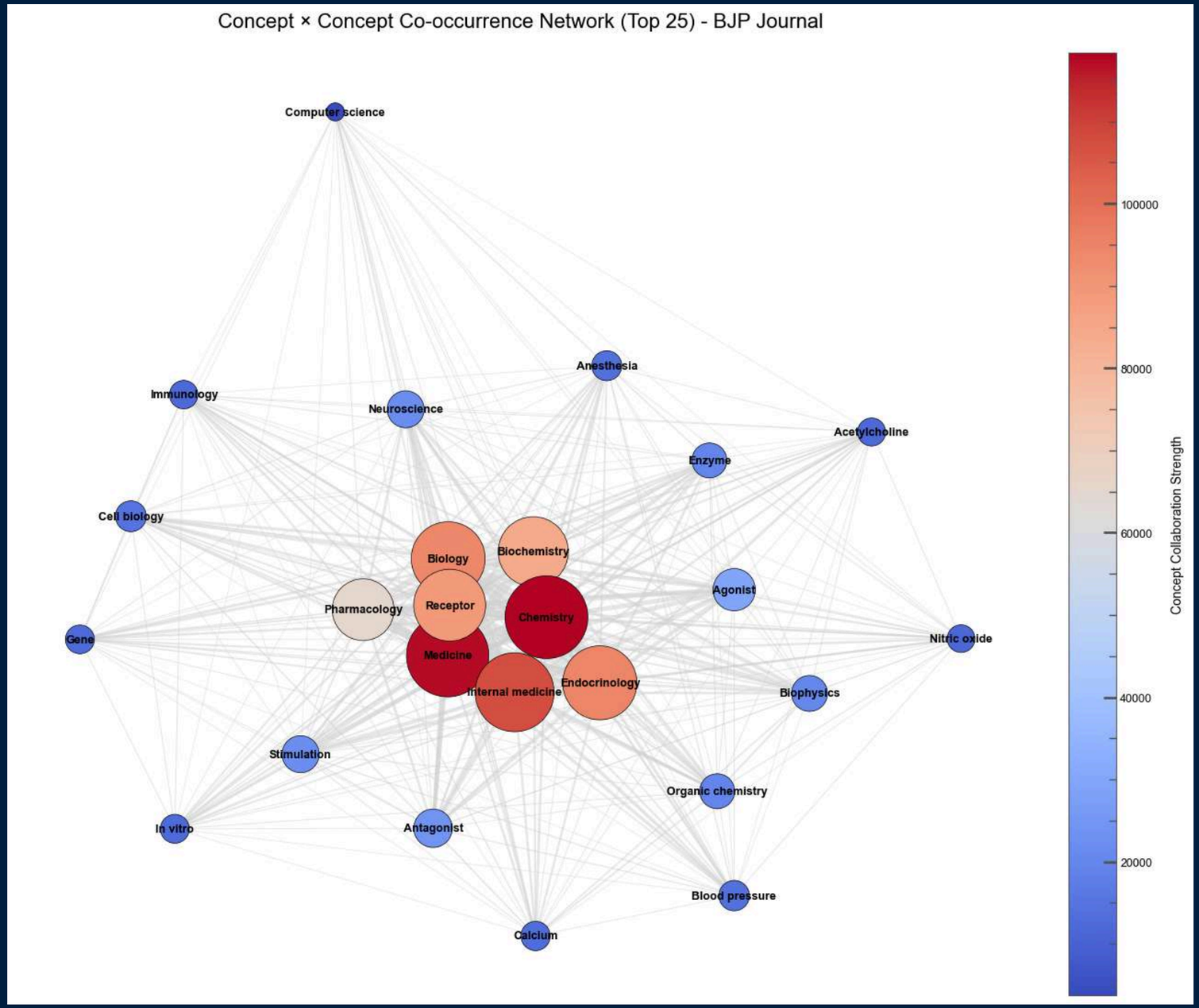


HSDSLab
Human & Social Data Science

Topics

Top Concepts- Co-Occurrence Network - BJP

- Selected the top 25 concepts by total number of collaborations.
- Co-Occurrence = two concepts co-occuring in the same article; counted once per article.

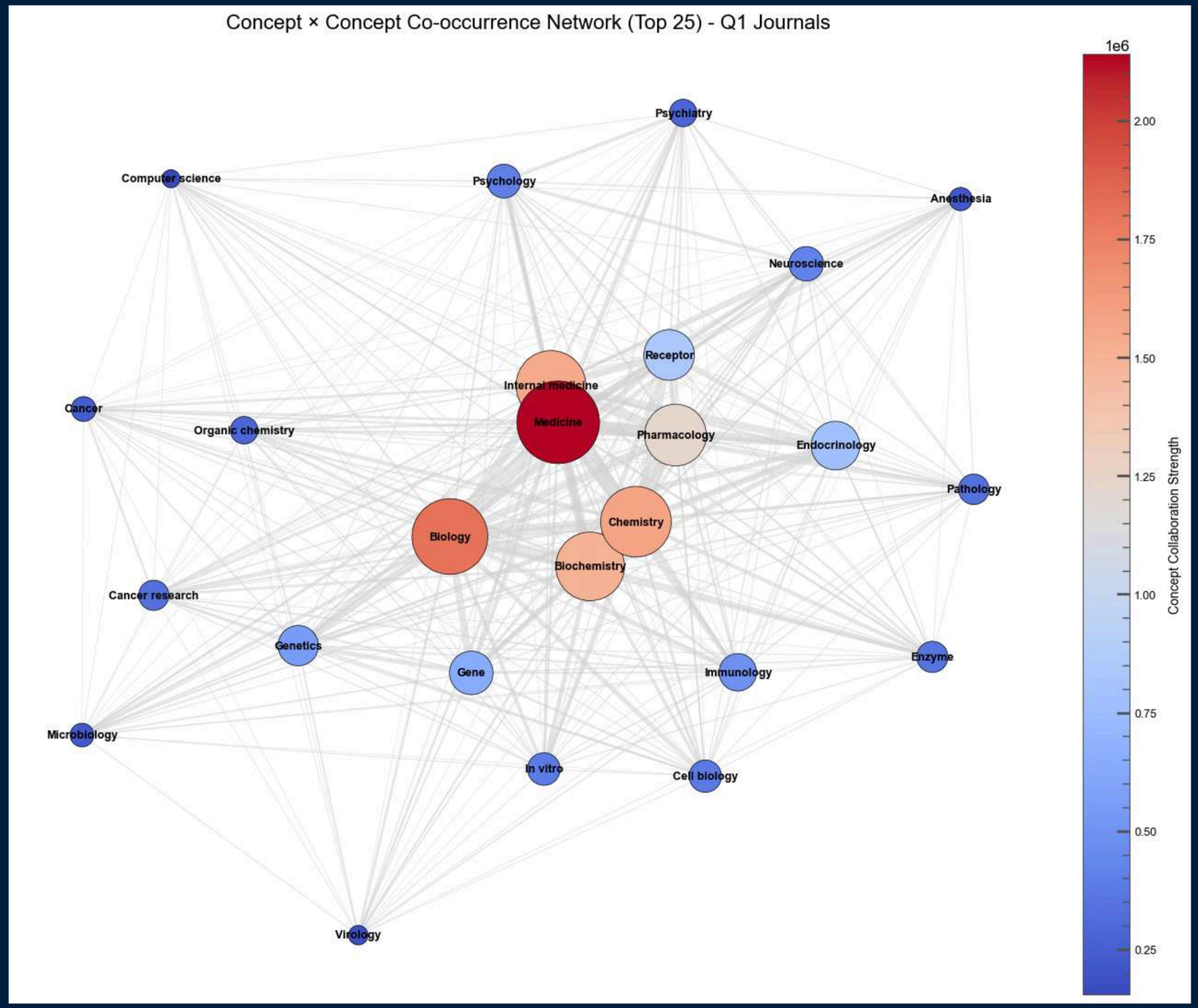


HSDSLab
Human & Social Data Science

Topics

Top Concepts - Co-Occurrence Network - Q1

- Selected the top 25 concepts by total number of collaborations.
- Co-Occurrence = two concepts co-occuring in the same article; counted once per article.

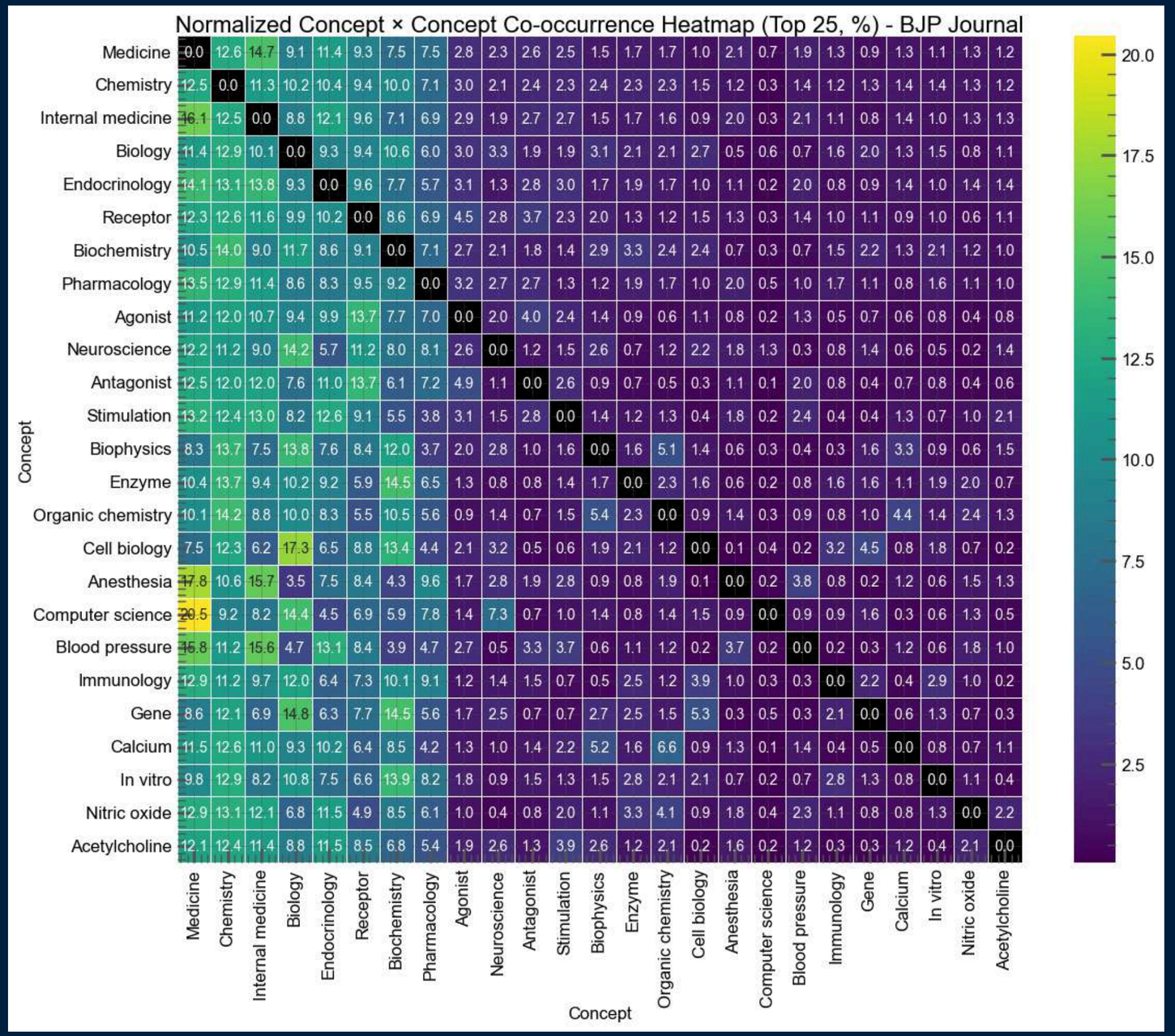


HSDSLab
Human & Social Data Science

Topics

Top Concepts - BJP Co-Occurrence Heatmap

- Rows = each concept's distribution profile → How this concept co-occurs with others concepts across articles (Each number : % of the concept's collabs)
- Columns = each concept's importance as a partner → How much other concepts are associated with this concept.

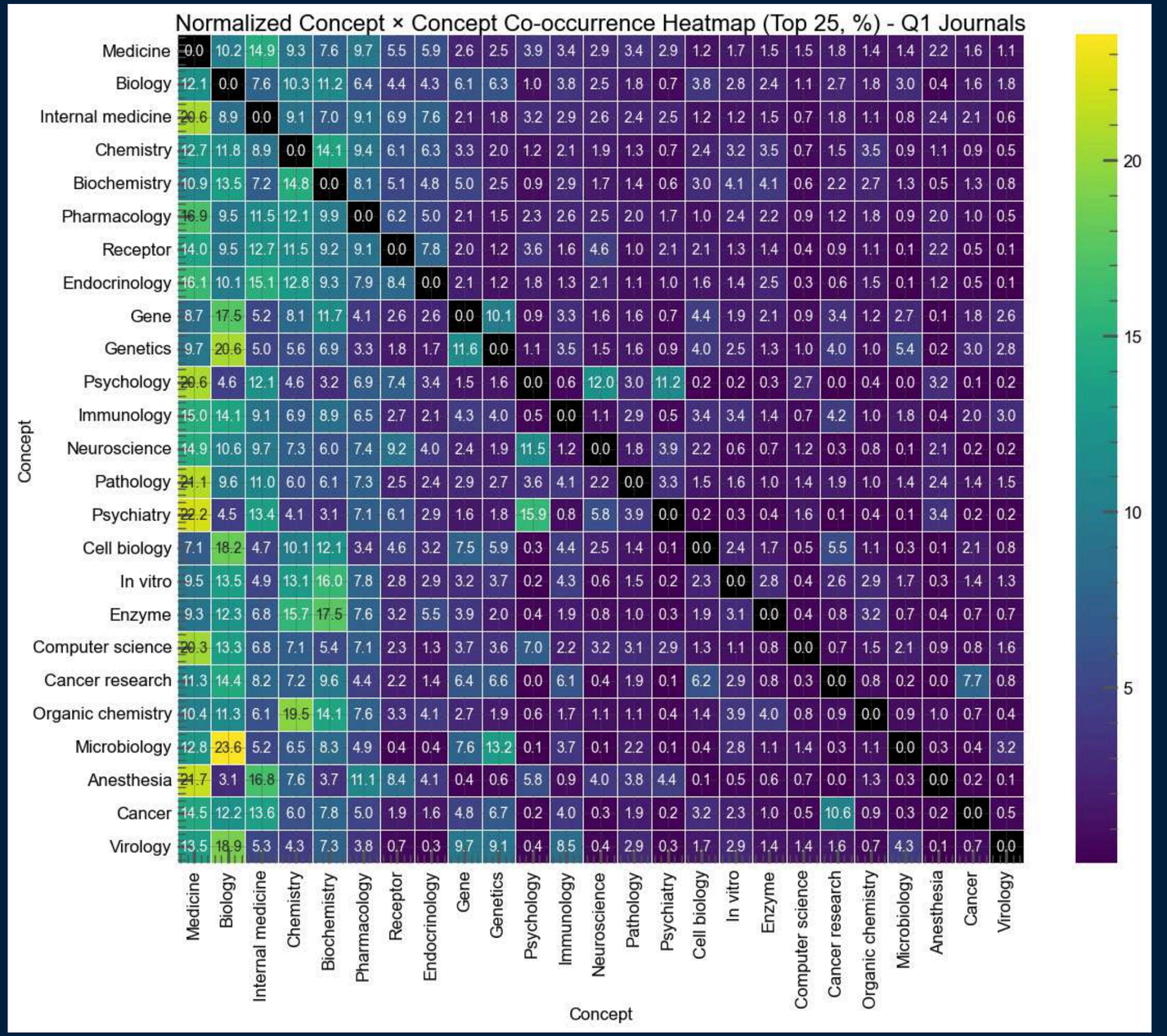


HSDSLab
Human & Social Data Science

Topics

Top Concepts - Q1 Co-Occurrence Heatmap

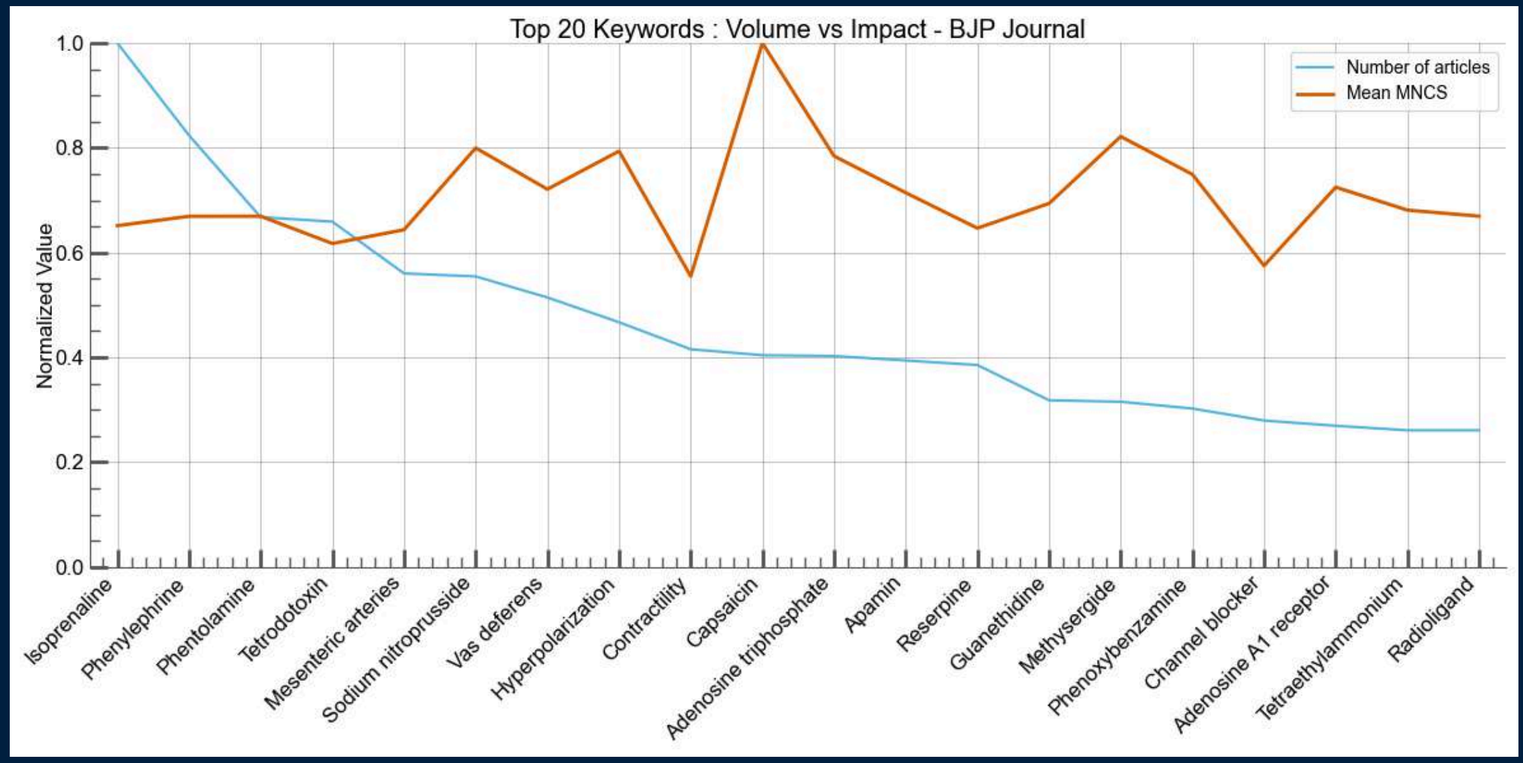
- Rows = each concept's distribution profile → How this concept co-occurs with others concepts across articles (Each number : % of the concept's collabs)
- Columns = each concept's importance as a partner → How much other concepts are associated with this concept.



HSDSLab
Human & Social Data Science

Topics

Top Keywords - BJP Publications vs MNCS



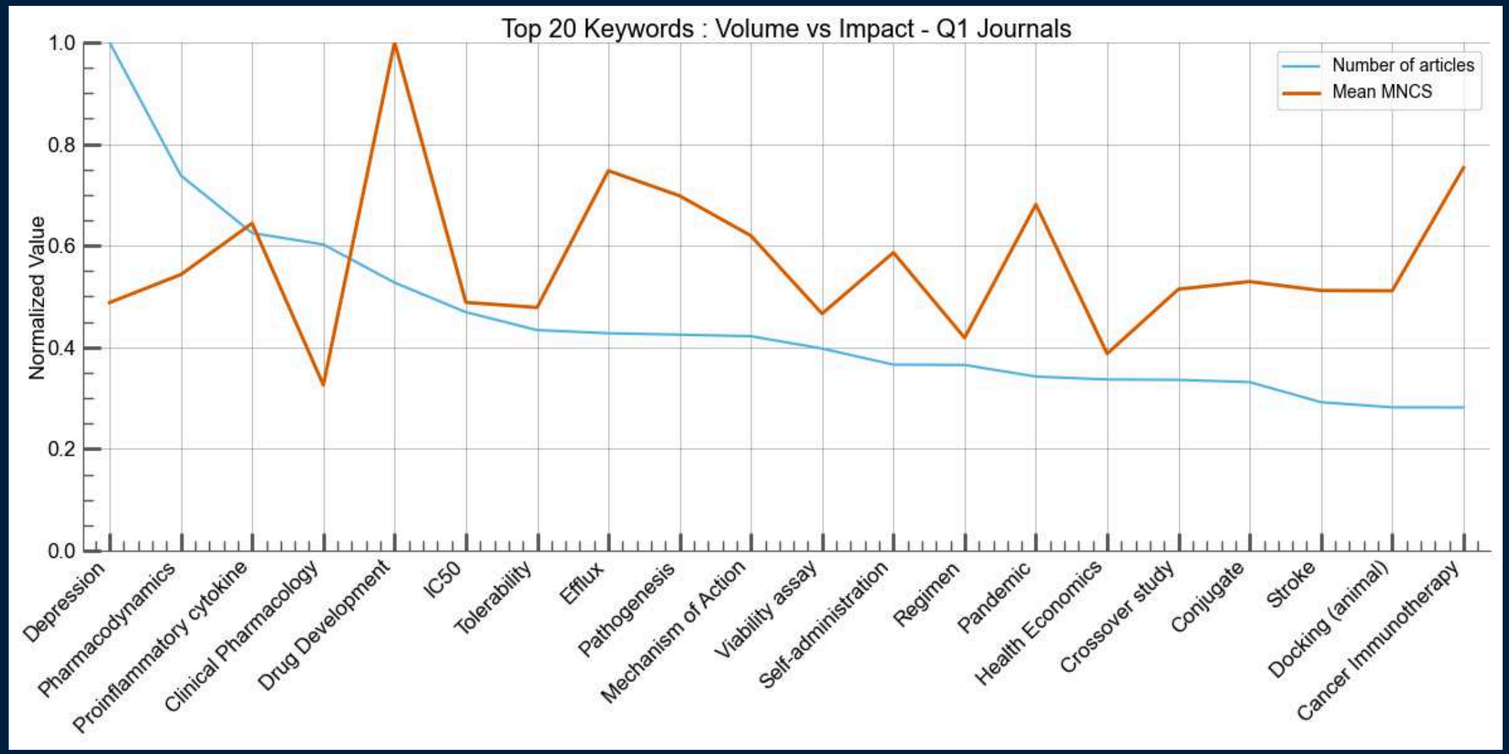
- Keywords are variables from OpenAlex.
- They represent specific terms or phrases describing the content of a publication.
- Each article can have multiple keywords, reflecting its main topics or methods.

HSDSLab
Human & Social Data Science

Topics

Top Keywords - Q1 Publications vs MNCS

- Keywords are variables from OpenAlex.
- They represent specific terms or phrases describing the content of a publication.
- Each article can have multiple keywords, reflecting its main topics or methods.

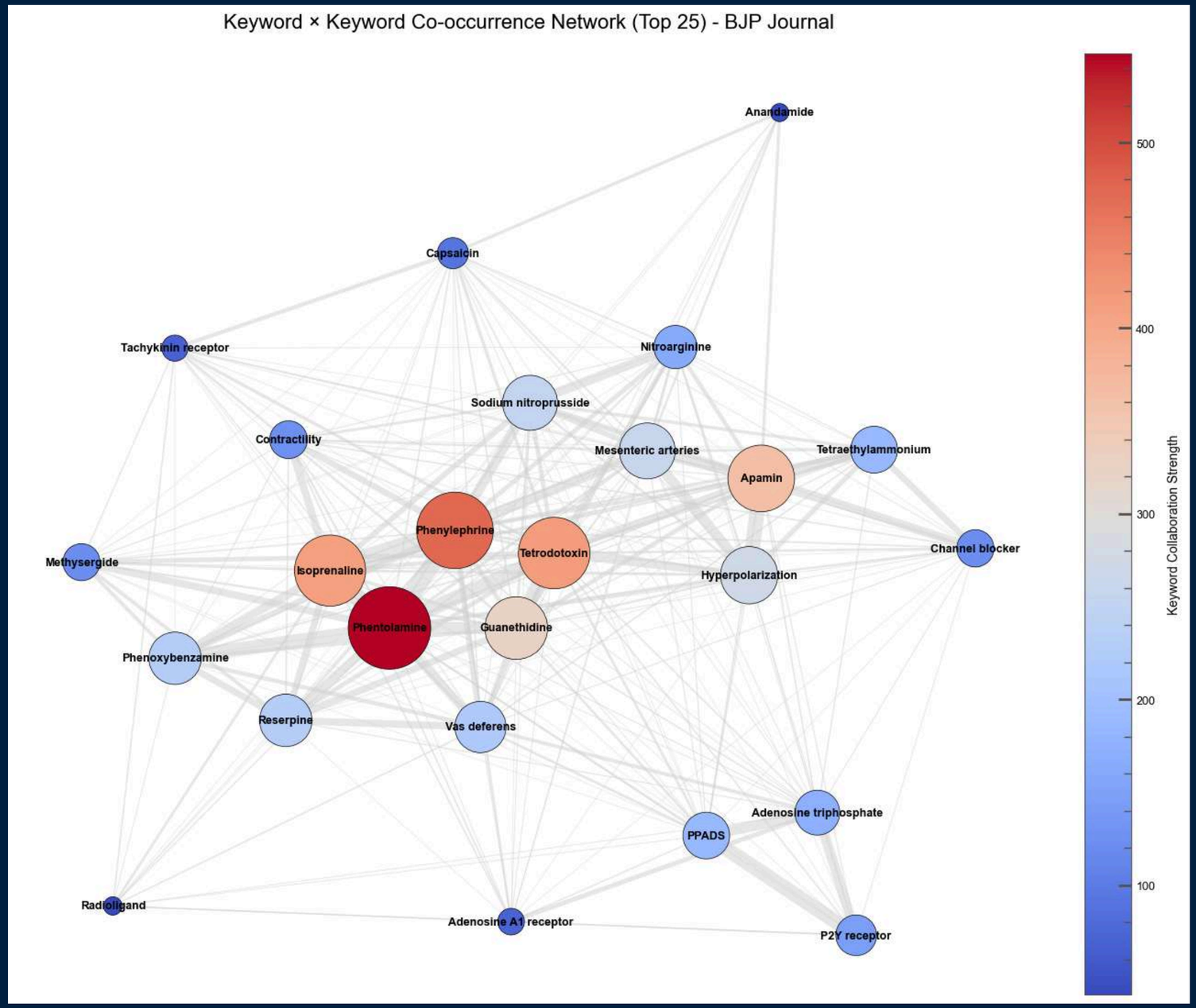


HSDSLab
Human & Social Data Science

Topics

Top Keywords - Co-Occurrence Network - BJP

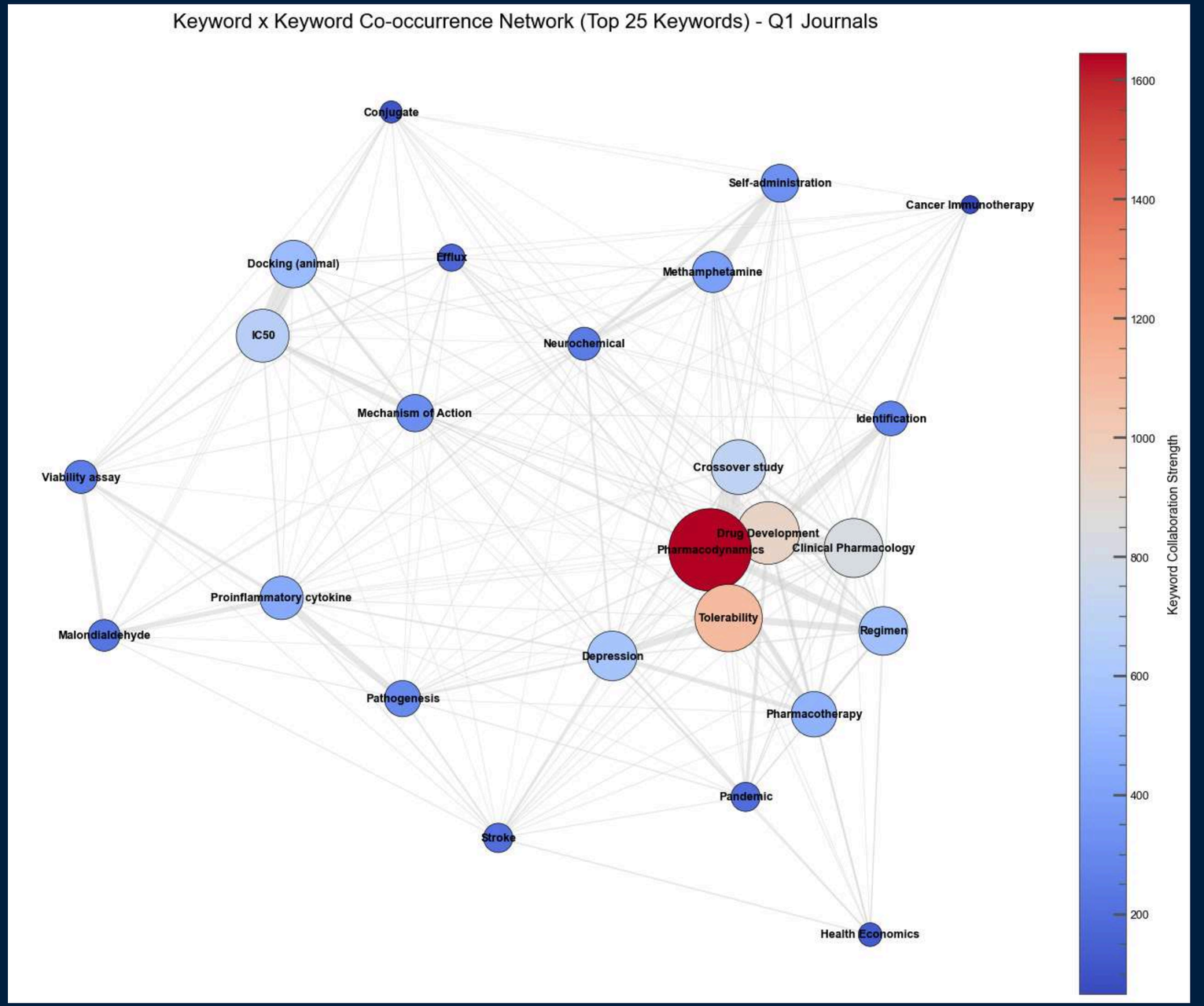
- Selected the top 25 keywords by total number of collaborations.
- Co-Occurrence = two keywords co-occuring in the same article; counted once per article.



Topics

Top Keywords - Co-Occurrence Network - Q1

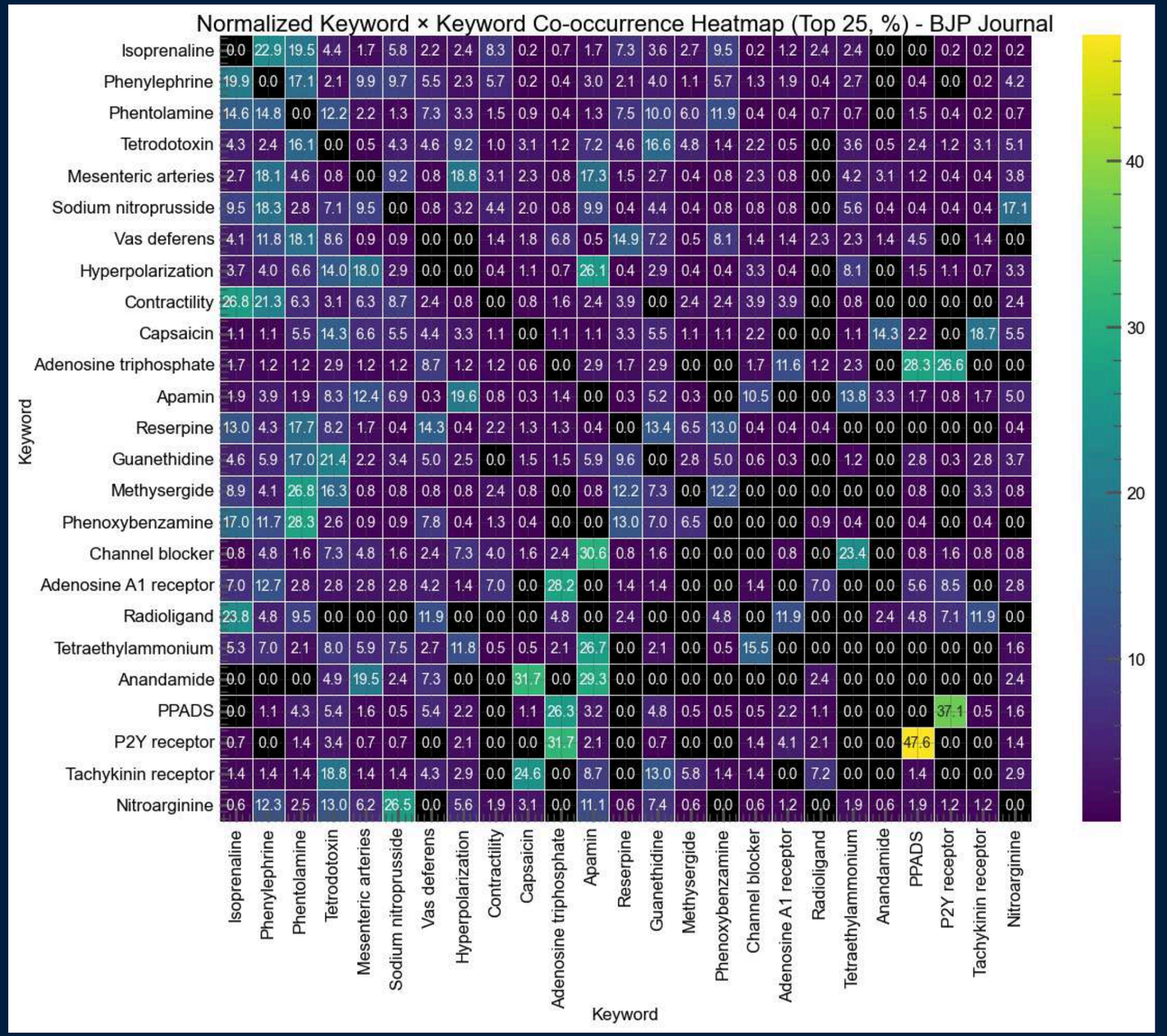
- Selected the top 25 keywords by total number of collaborations.
- Co-Occurrence = two keywords co-occuring in the same article; counted once per article.



Topics

Top Keywords - BJP Co-Occurrence Heatmap

- Rows = each keyword's distribution profile → How this keyword co-occurs with others keywords across articles (Each number : % of the keyword's collabs)
- Columns = each keyword's importance as a partner → How much other keywords are associated with this keyword.

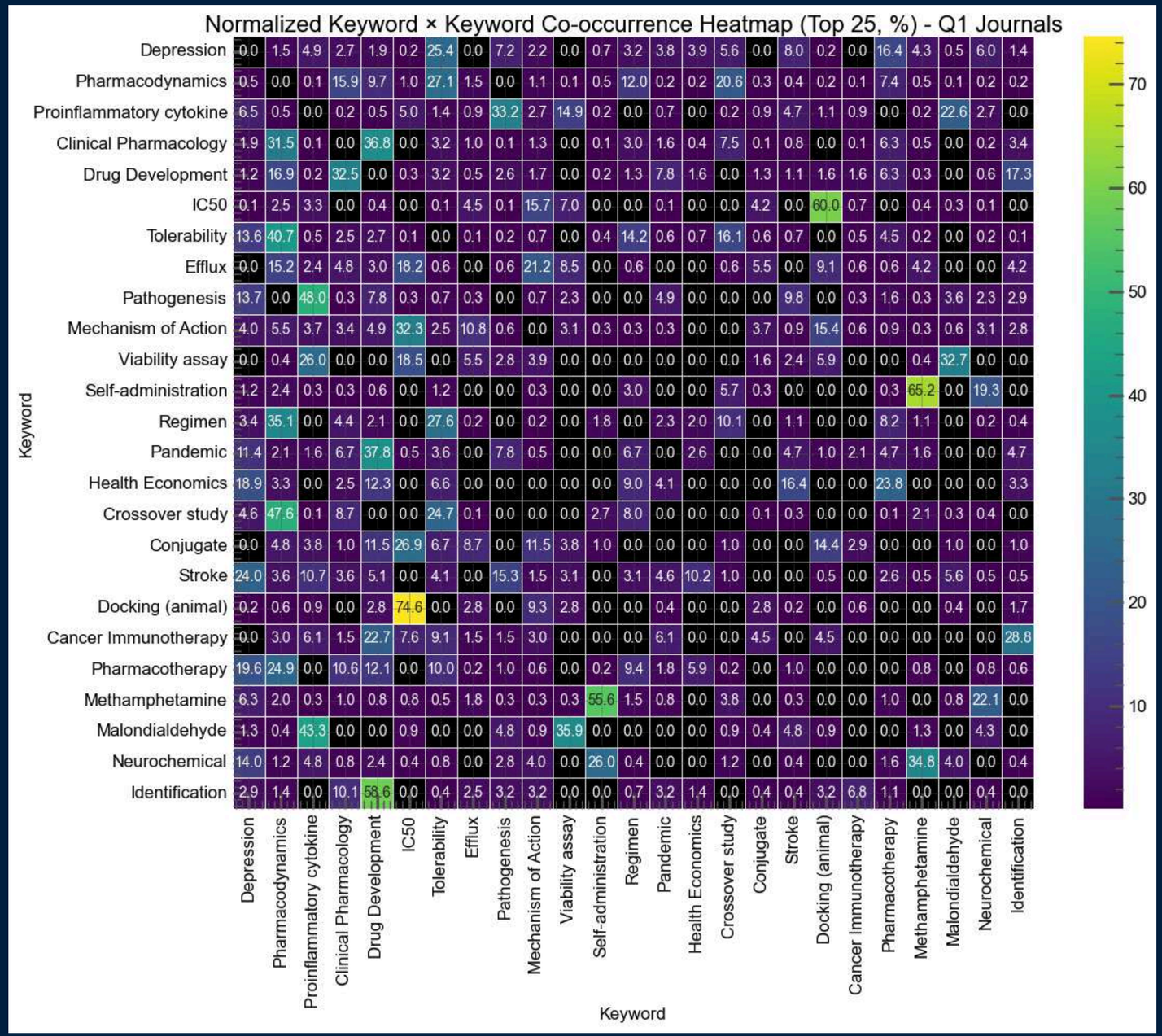


HSDSLab
Human & Social Data Science

Topics

Top Keywords - Q1 Co-Occurrence Heatmap

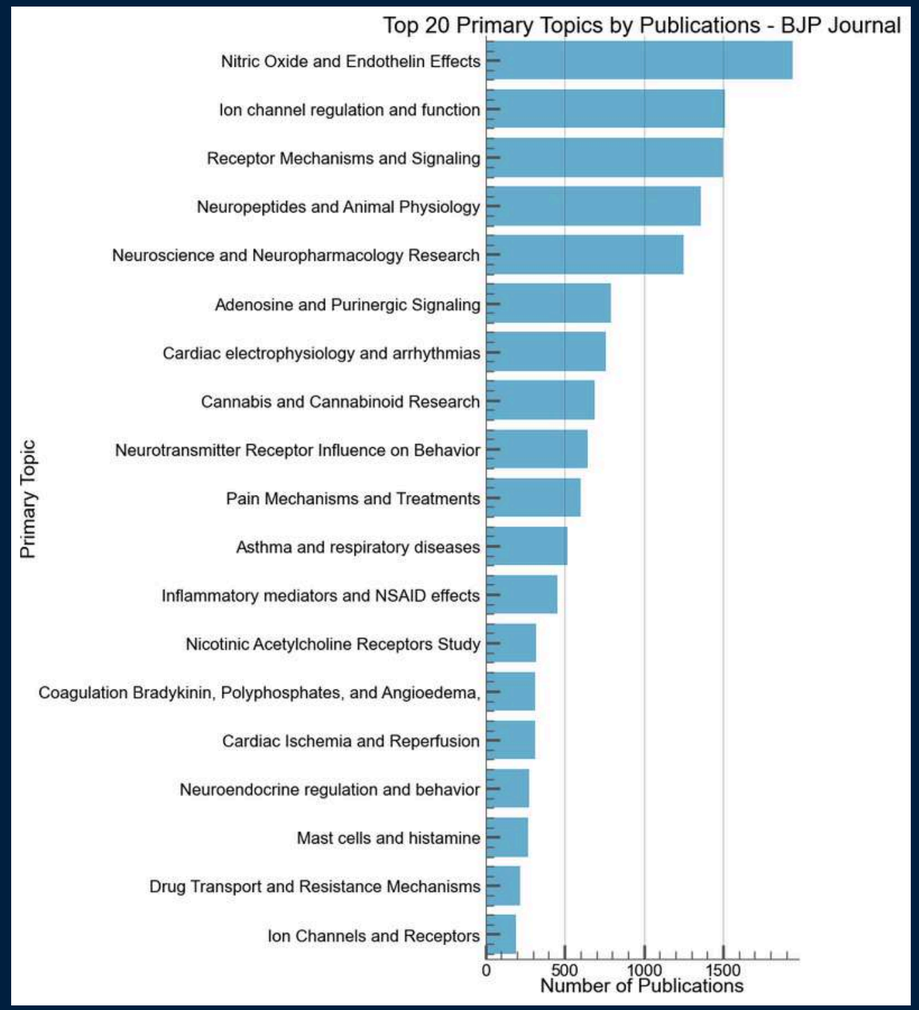
- Rows = each keyword's distribution profile → How this keyword co-occurs with others keywords across articles (Each number : % of the keyword's collabs)
- Columns = each keyword's importance as a partner → How much other keywords are associated with this keyword.

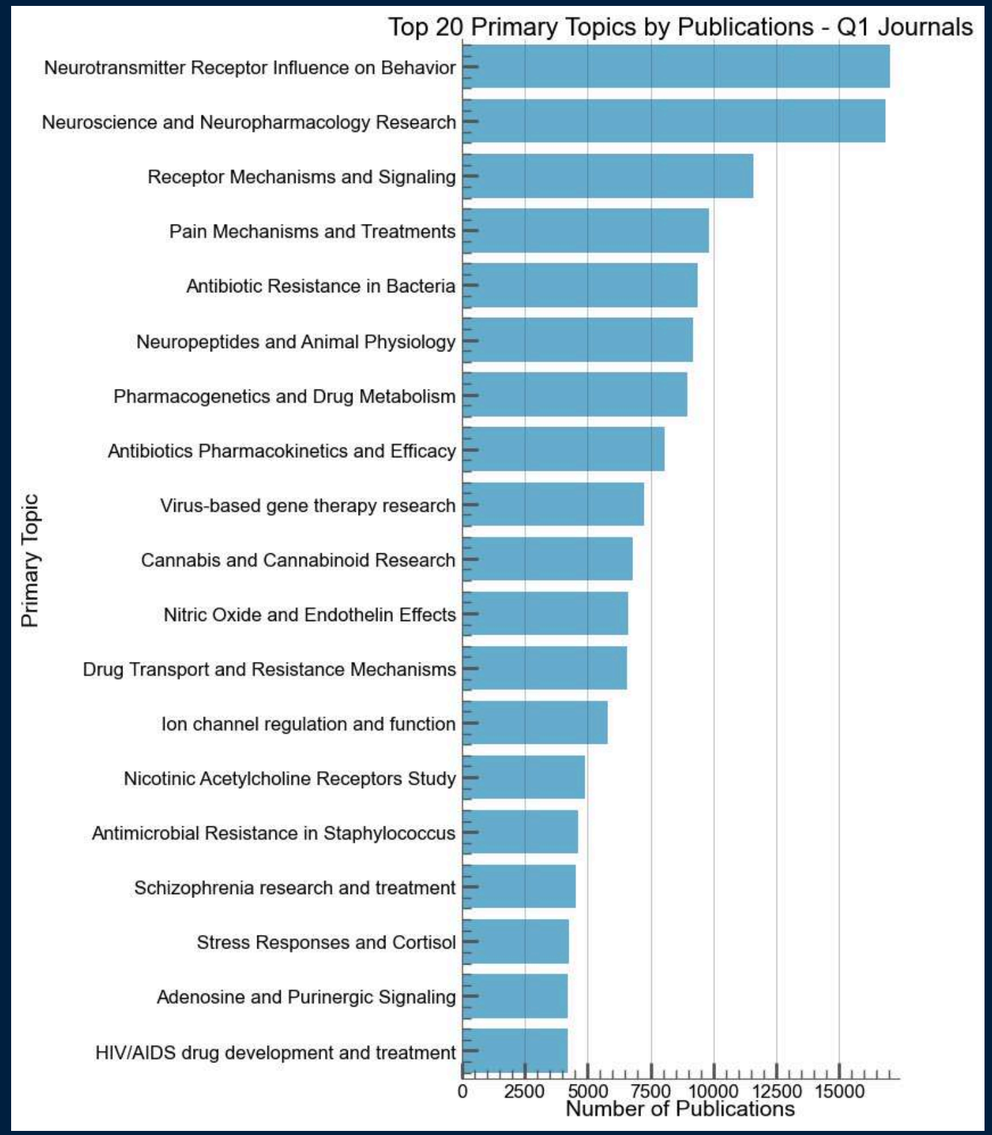


HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Publications

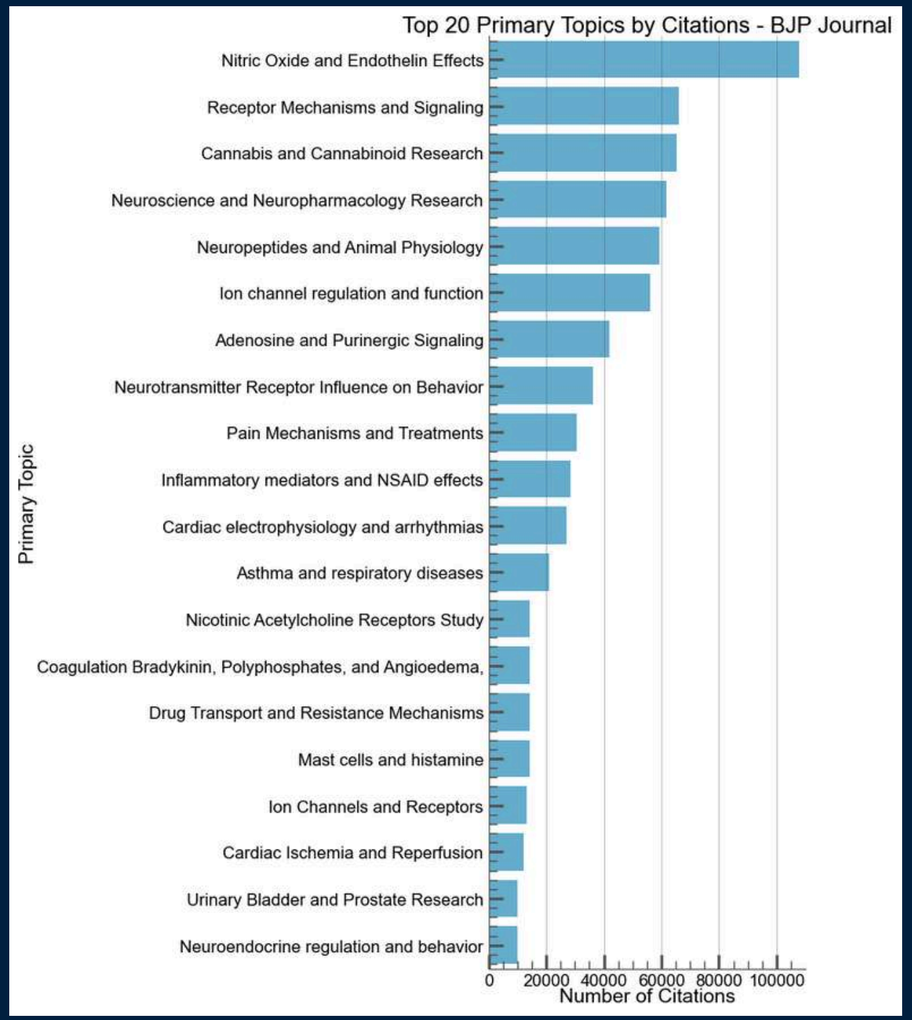


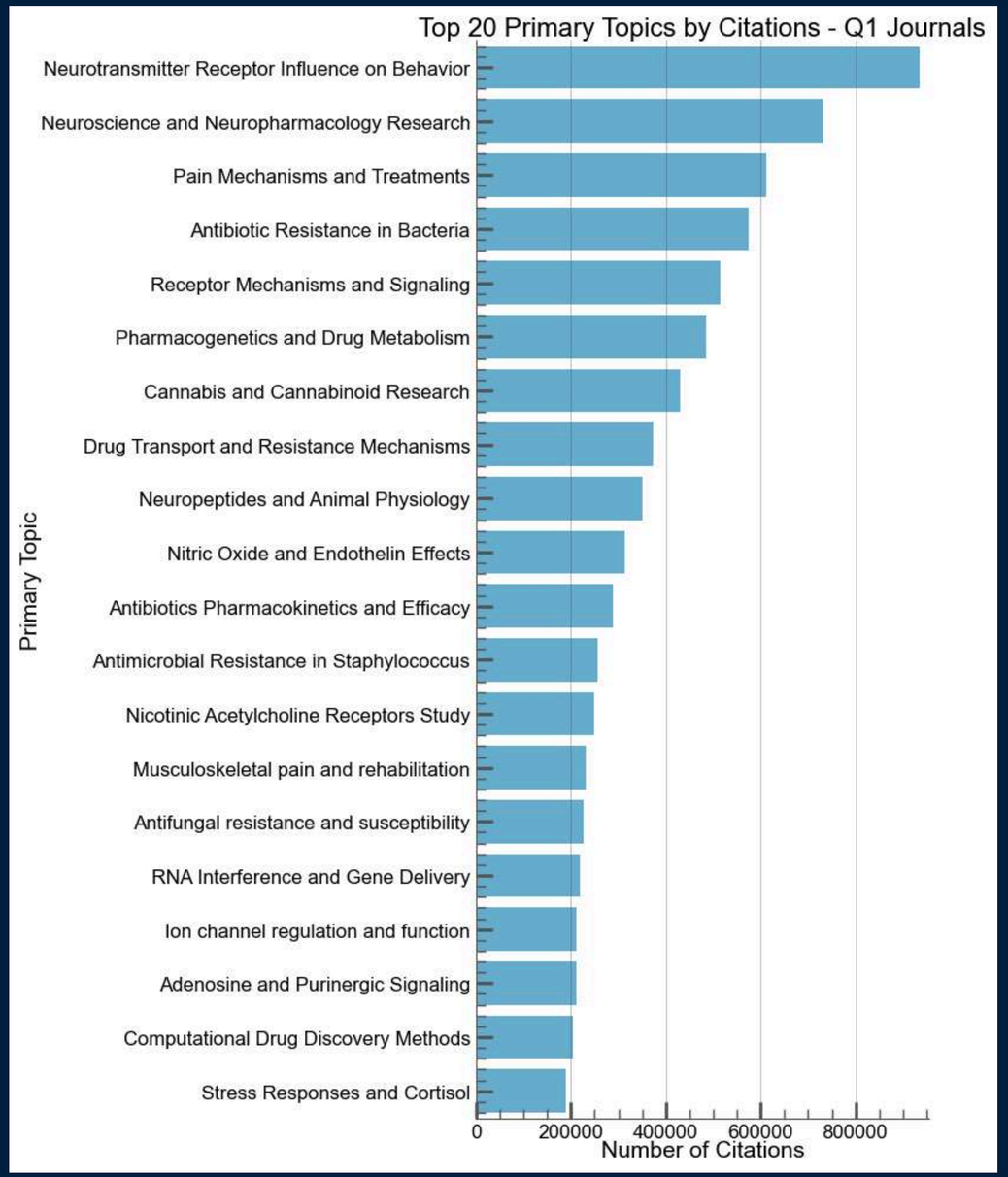


HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Citations



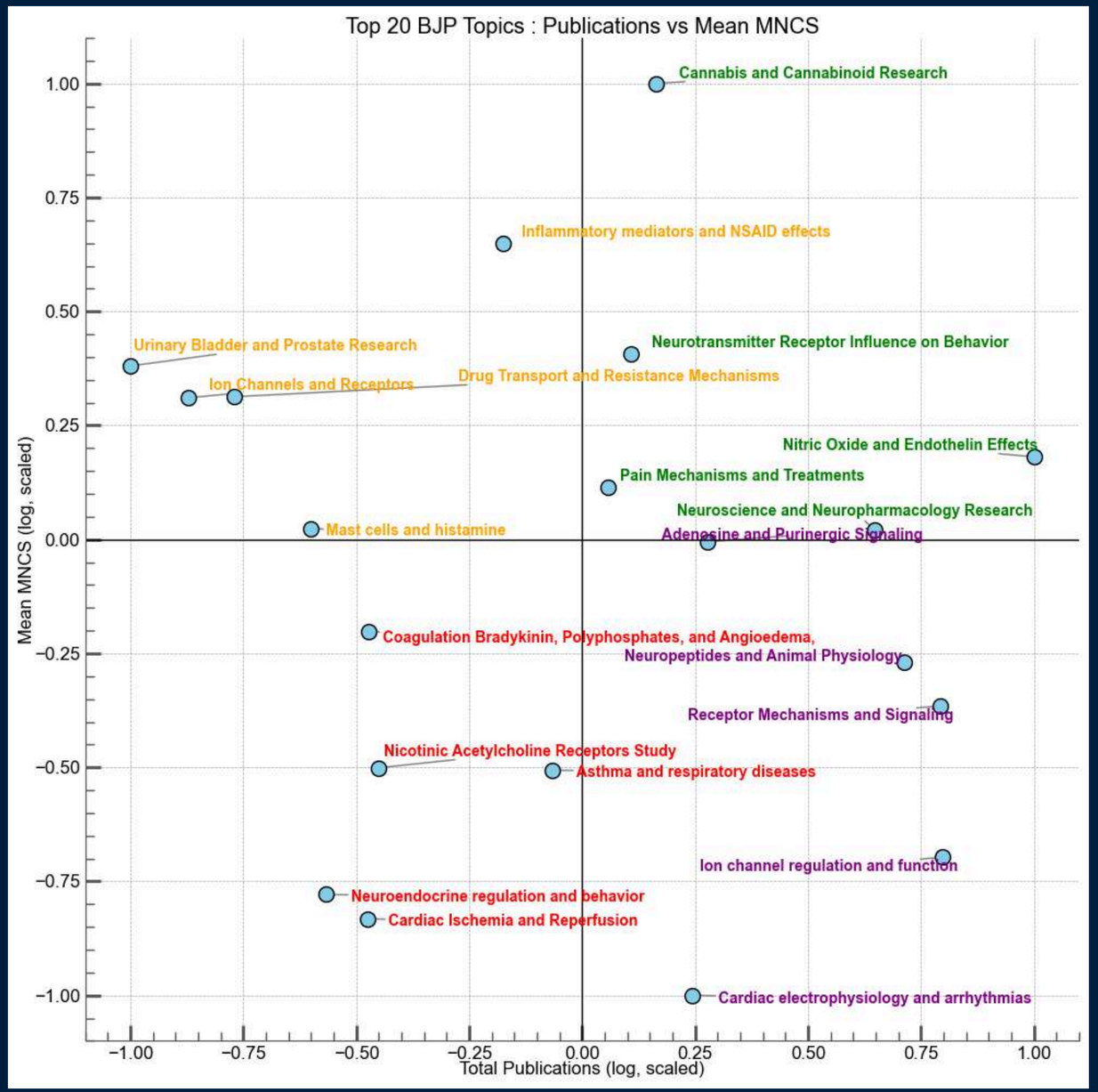


HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - BJP Publications vs Citations

- Topics with more publications tend to have a lower MNCS compared to other top topics
- Small underperformances : (*Cardiac Ischemia and reperfusion*, *Neuroendocrine regulation and behavior*)
- High-Impact : (*Cannabis and Cannabinoid Research, Nitric Oxide and Endothelin Effects)*

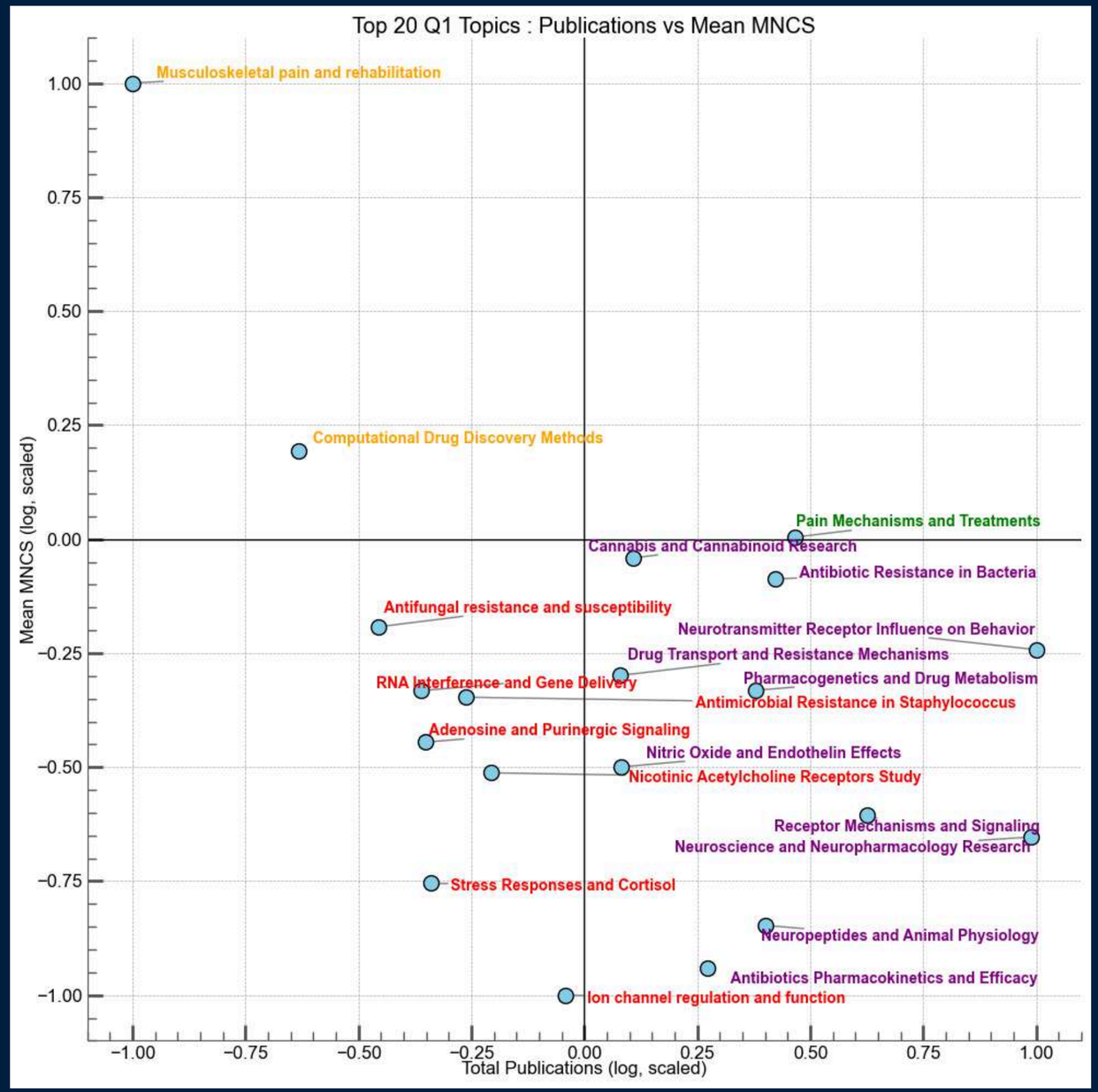


HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Q1 Publications vs Citations

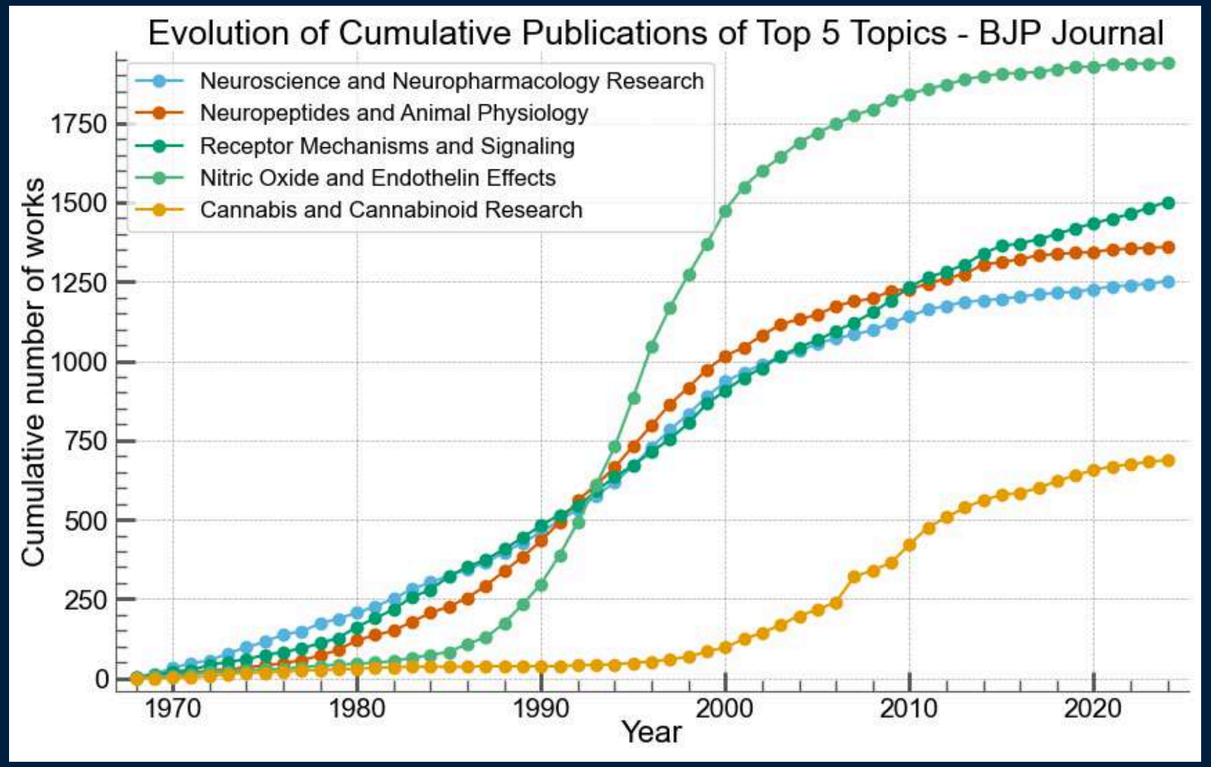
- Topics with more publications tend to have a lower MNCS compared to other top topics
- Small underperformances : (*Stress Responses and Cortisol*, ...)
- High-Impact : (*Pain Mechanisms and Treatments, ...)*

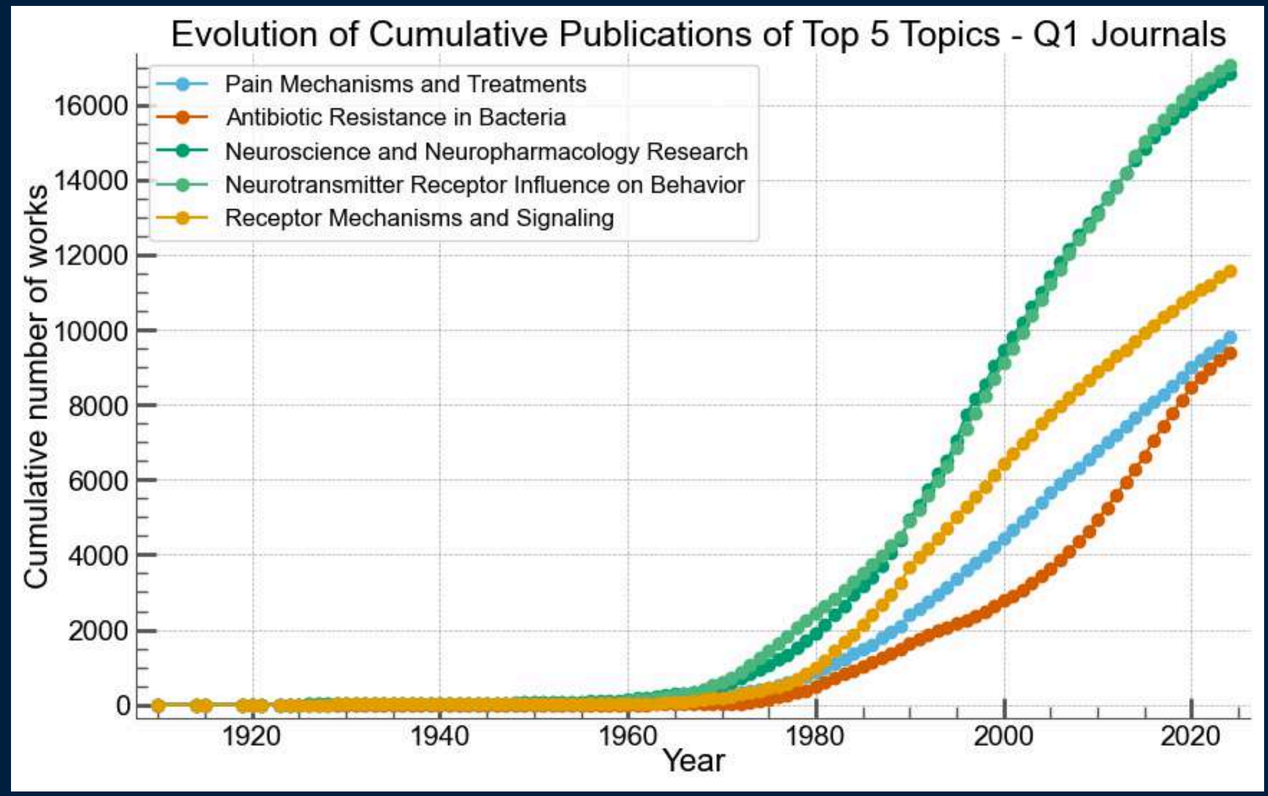


HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric -
Top 5 Topics Publications Evolution

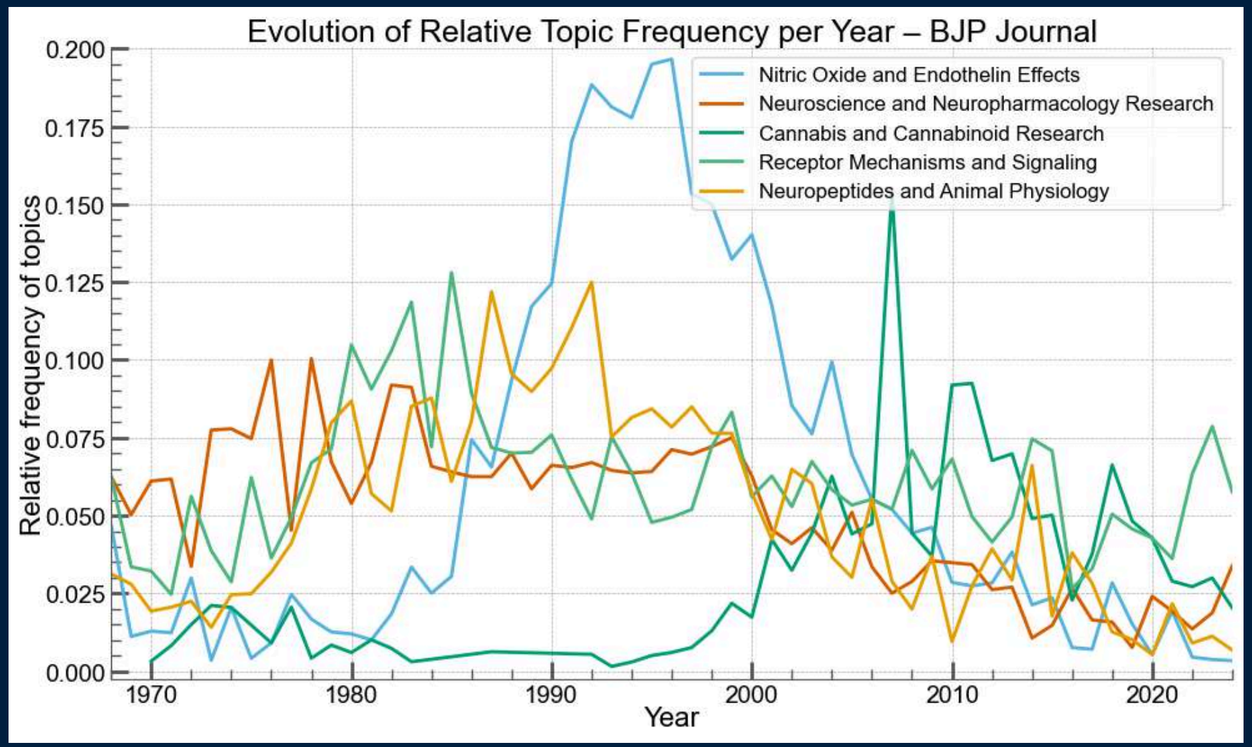


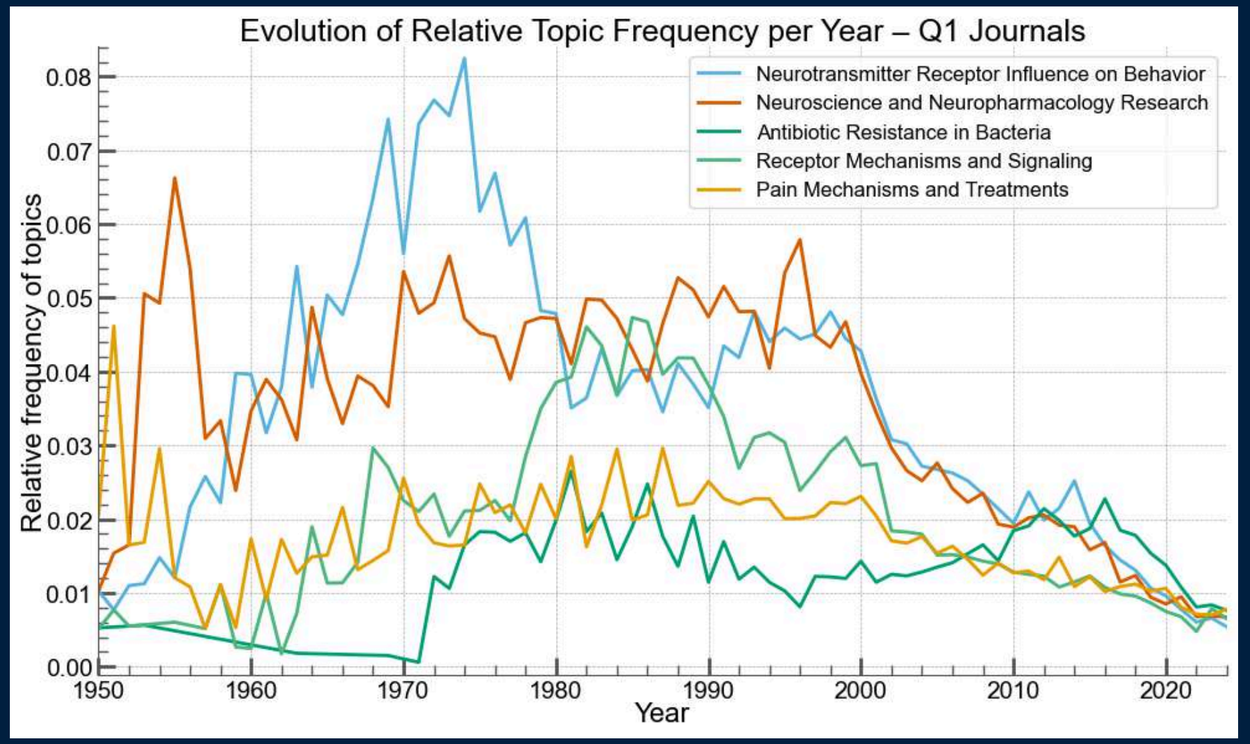


HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric -
Top 5 Topics *Frequency* Evolution





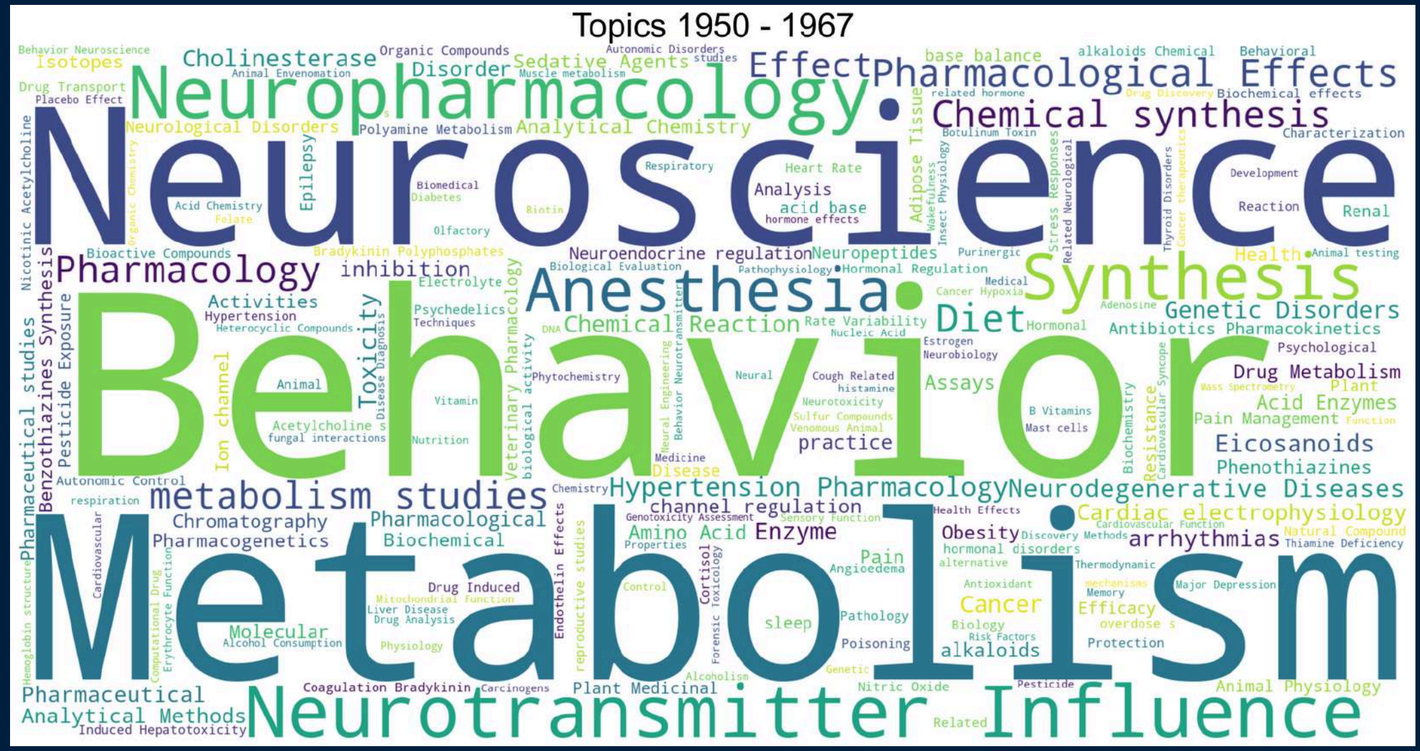
Frequency : Share of publications on a given topic relative to all publications in the same year.

HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Topics
Evolution through decades (Publications)

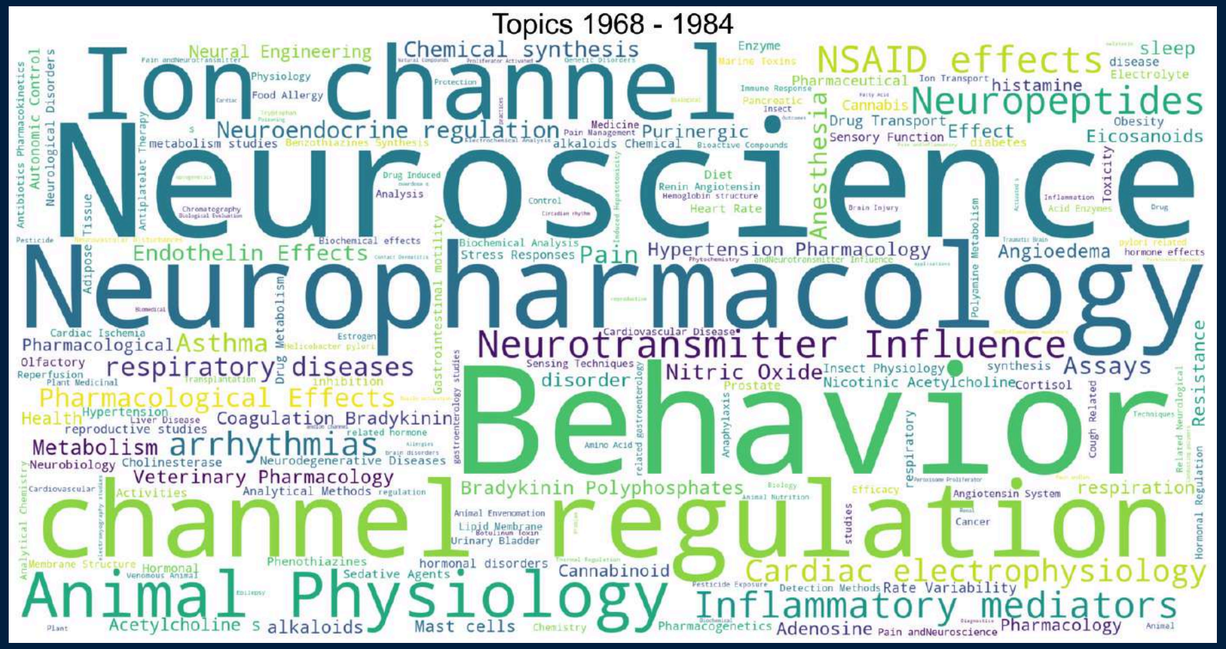
- Used WordCloud library from Python
- Q1 Journals for this period (no/few works detected from BJP)



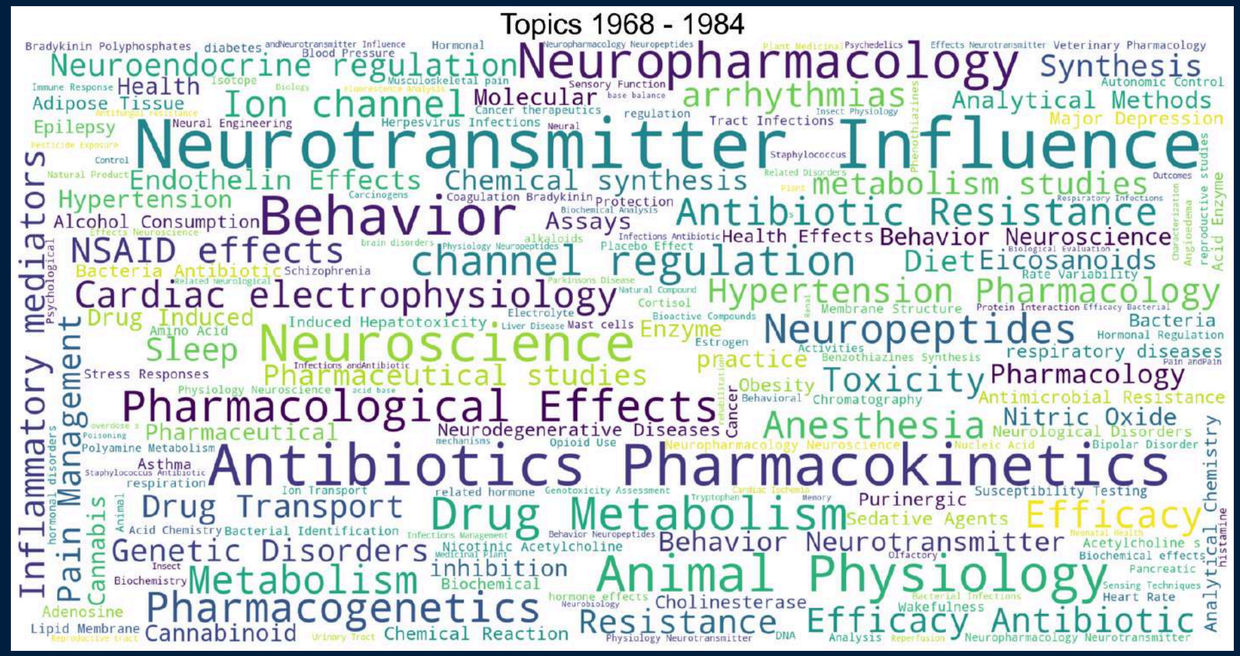
HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Topics
Evolution through decades (Publications)



BJP Journal

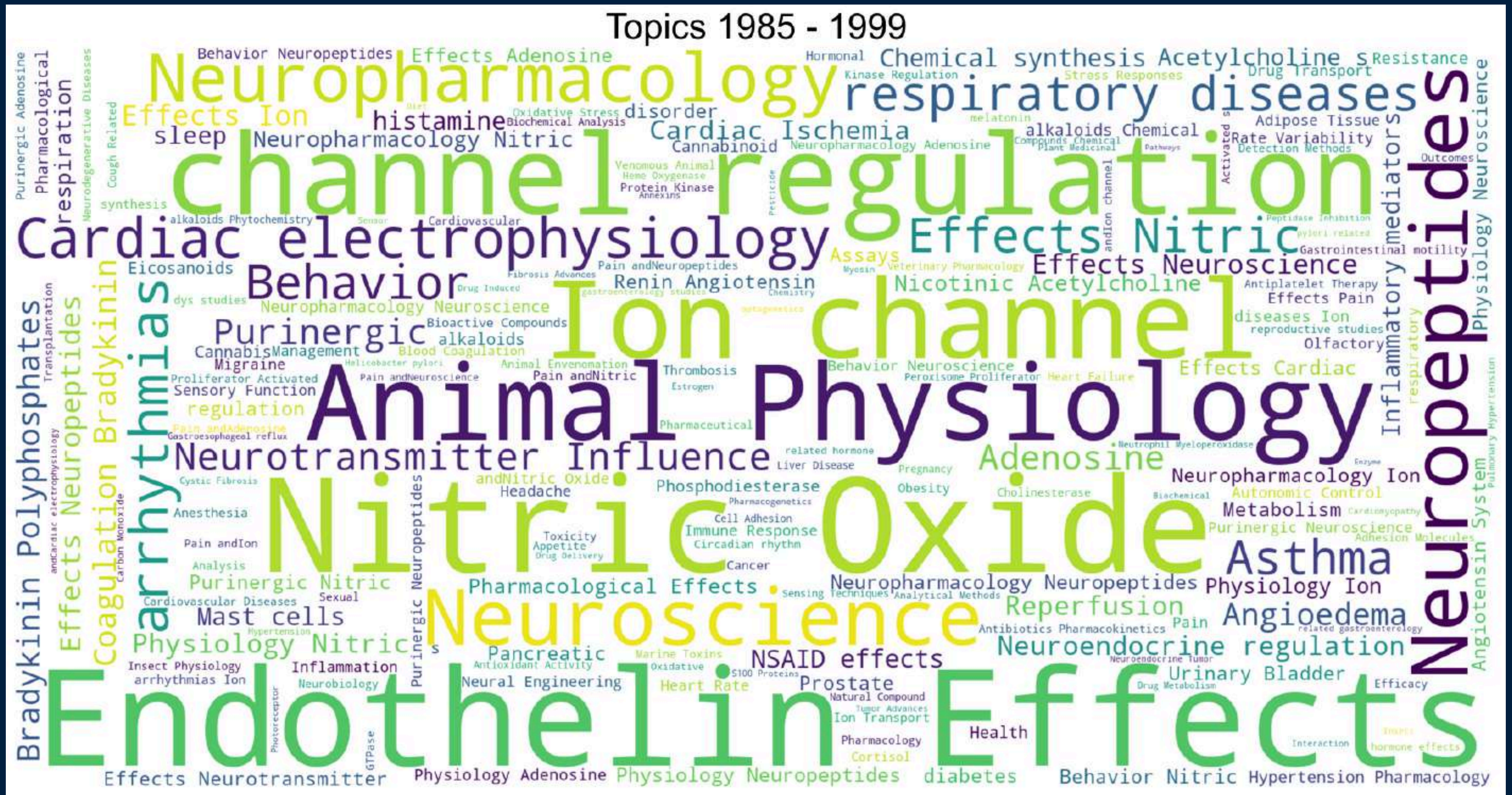


Q1 Journals

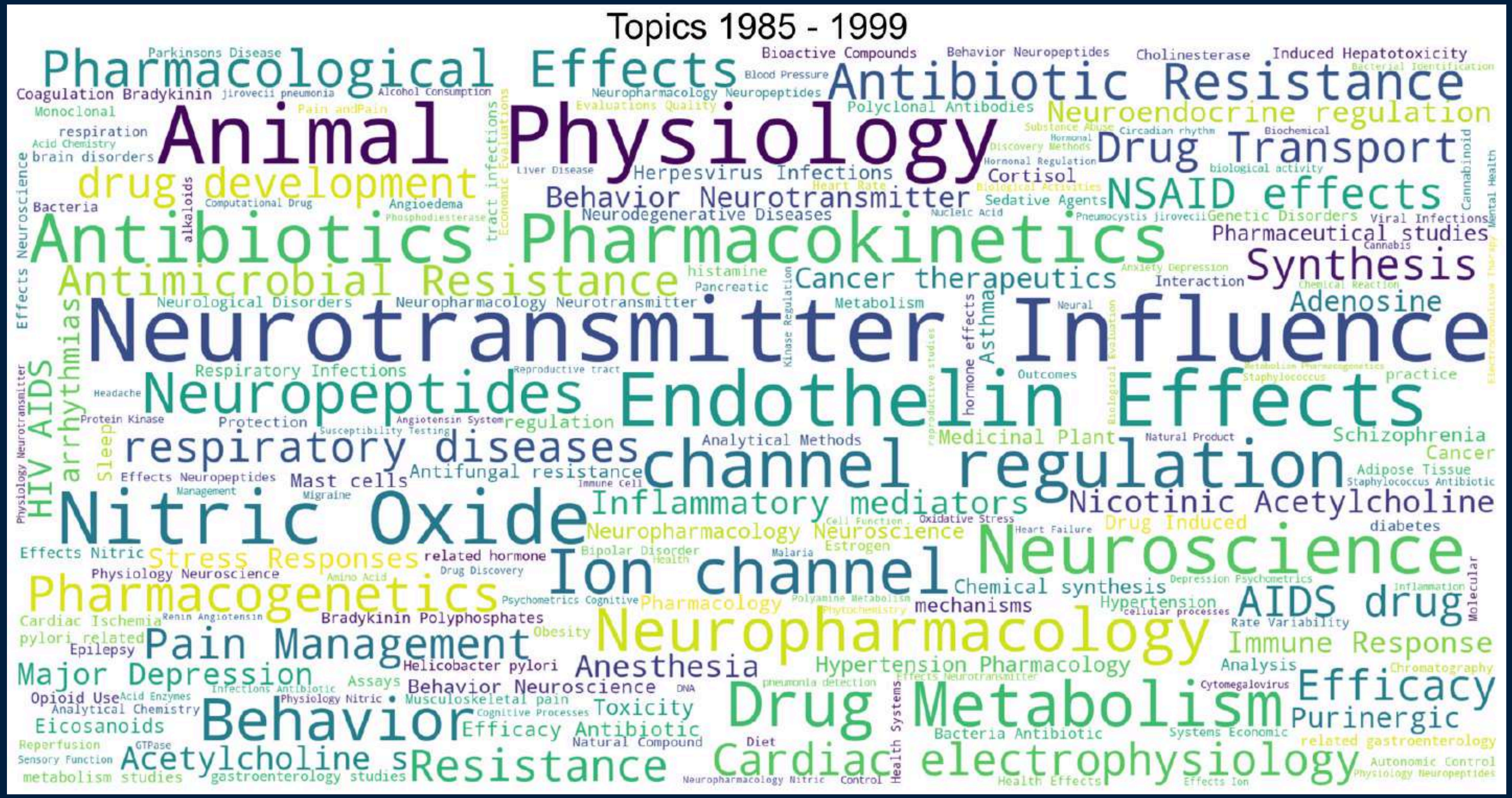
HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Topics
Evolution through decades (Publications)



BJP Journal

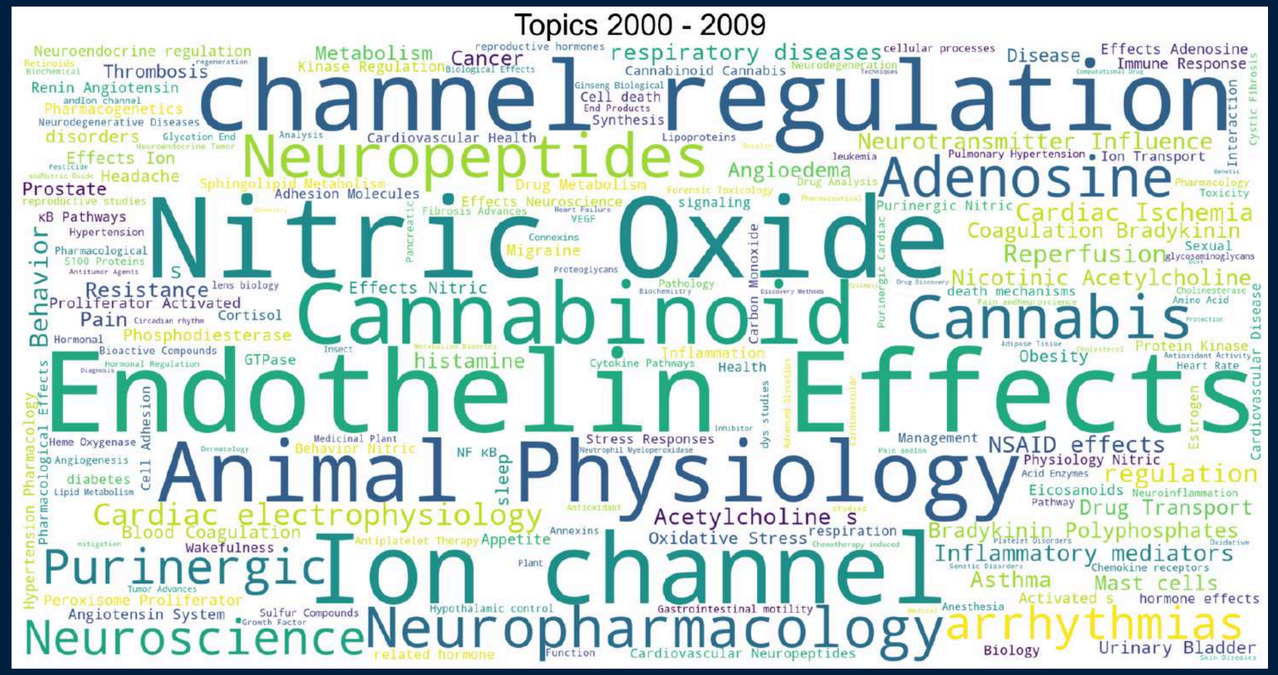


Q1 Journals

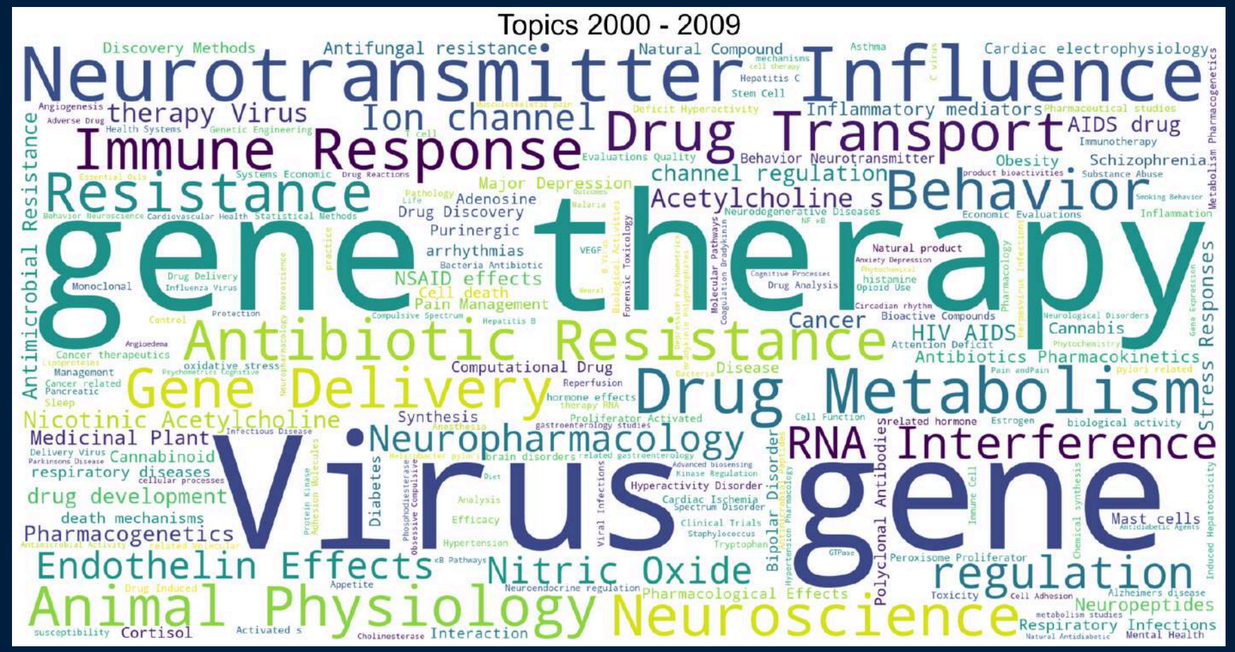
HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Topics
Evolution through decades (Publications)



BJP Journal

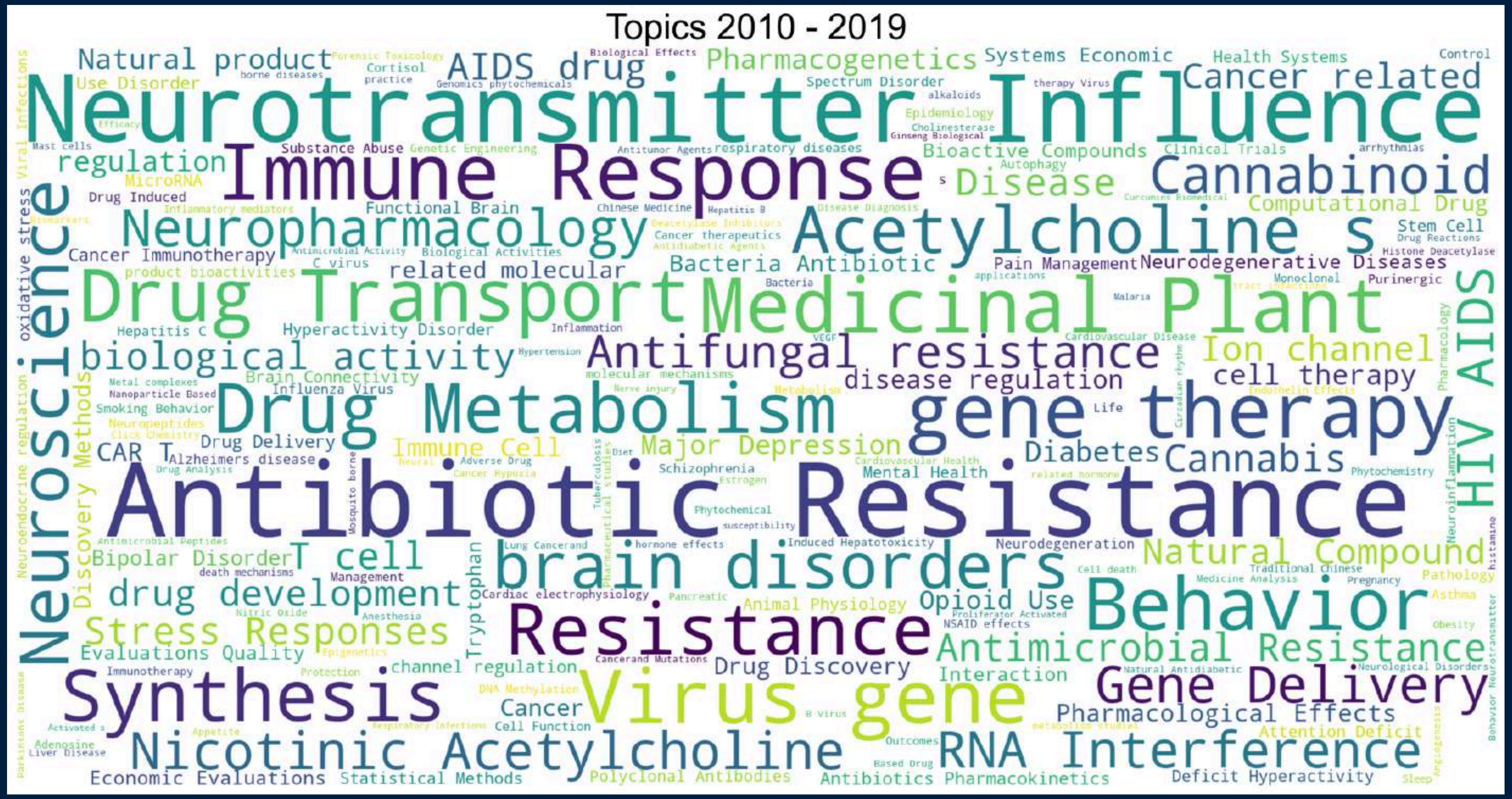


Q1 Journals

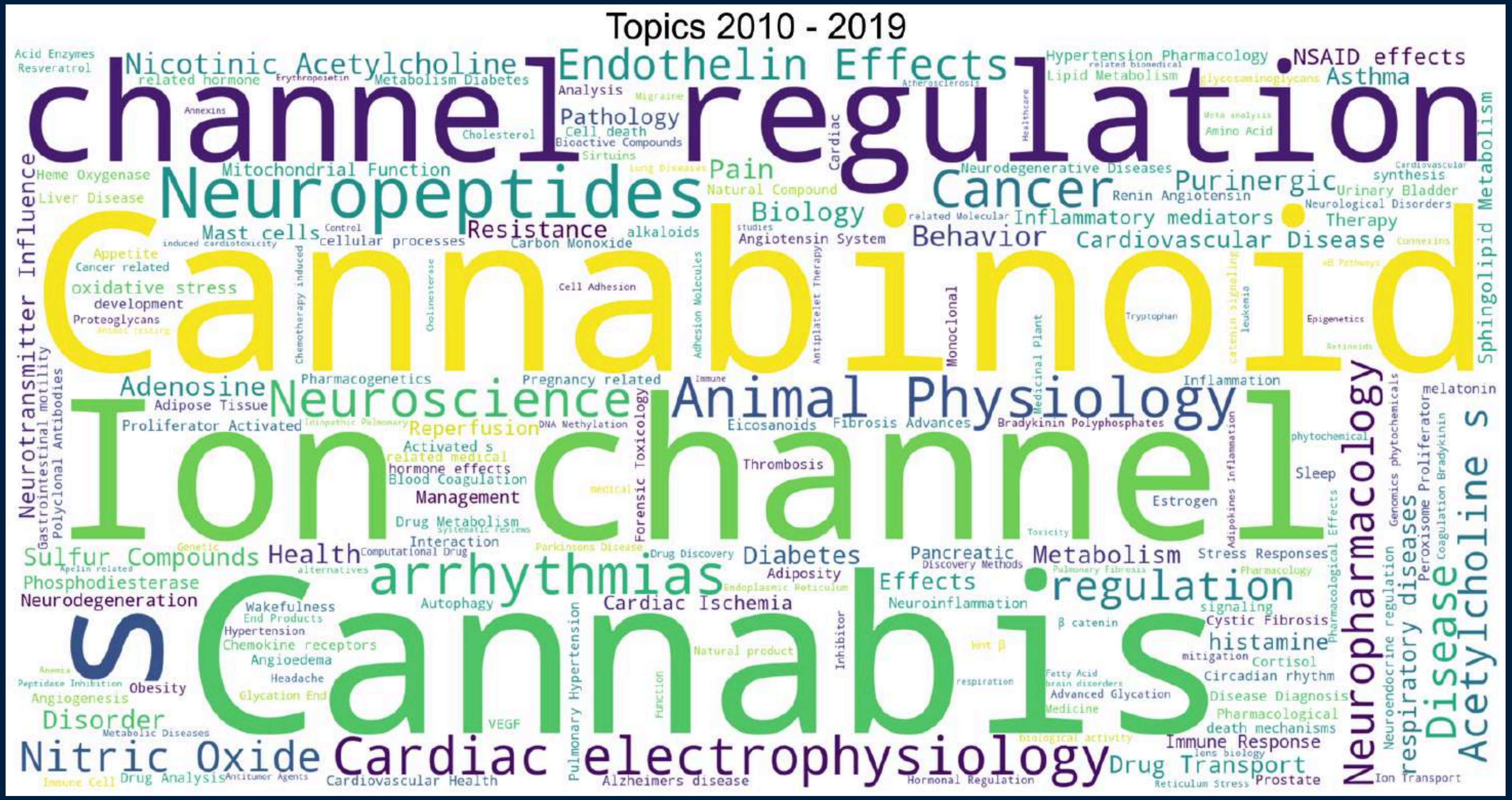
HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Topics
Evolution through decades (Publications)



Q1 Journals

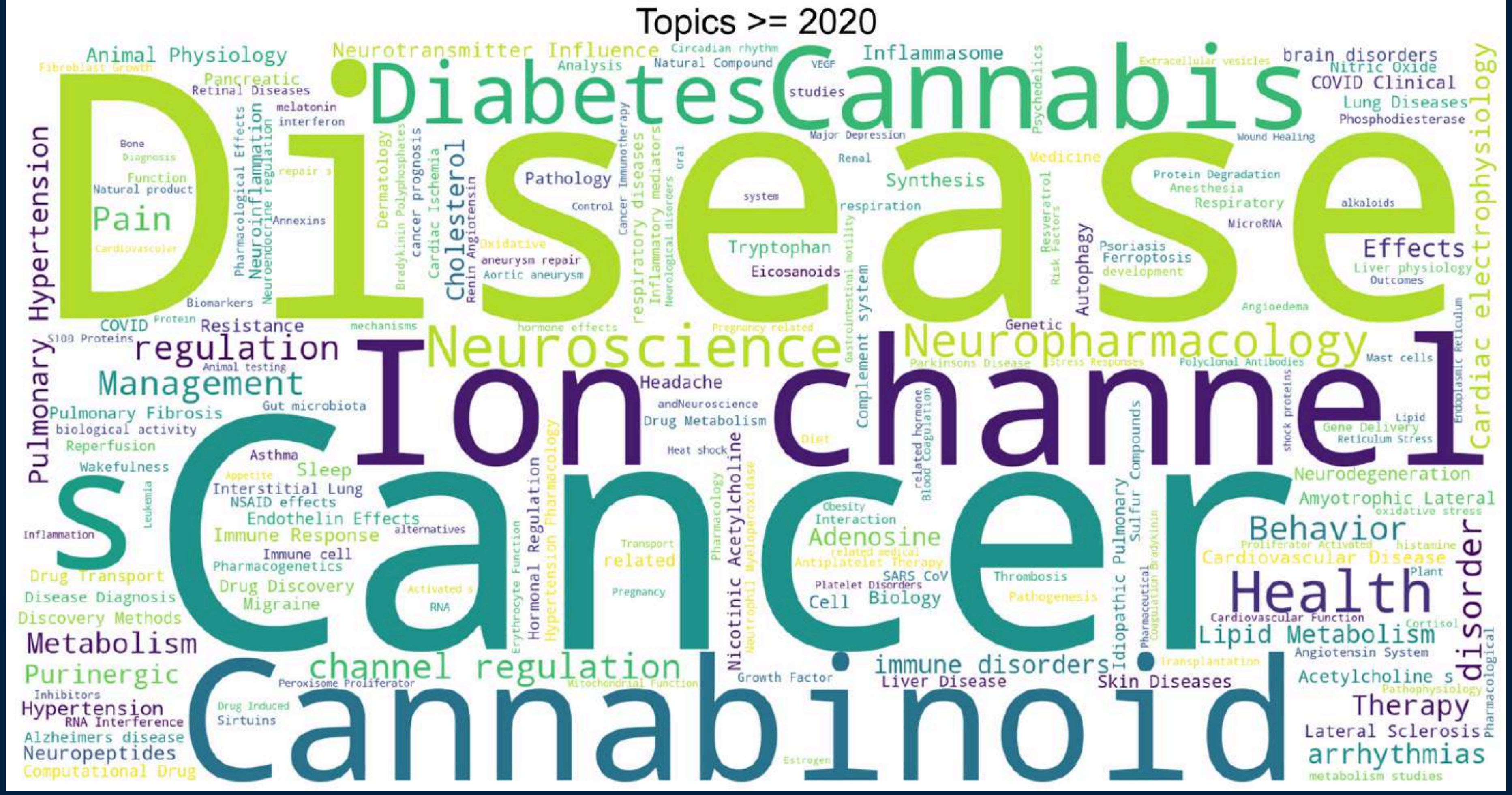


BJP Journal

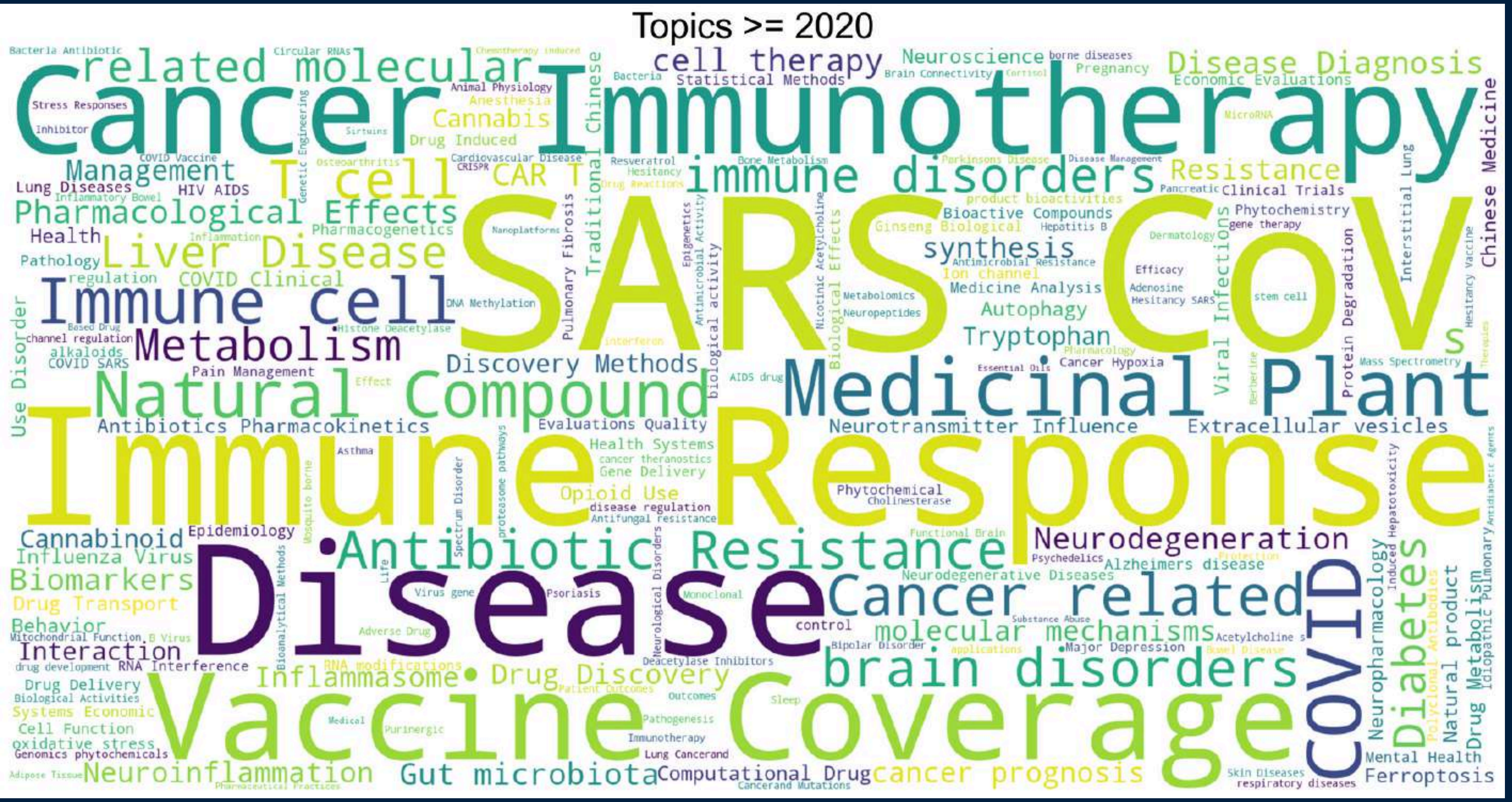
HSDSLab
Human & Social Data Science

Topics

The Primary Topic metric - Top Topics
Evolution through decades (Publications)



BJP Journal



Q1 Journals

Topics

Topic Modelling - General Presentation

BERTopic: Discovering Themes in 700k Scientific Works

- Groups similar papers into topics automatically using AI embeddings
- Clusters papers with UMAP + HDBSCAN and extracts keywords
- Assigns each paper to a single main topic with confidence score

Key advantage vs OpenAlex primary_topic :

- OpenAlex: 3077 manually assigned topics : very fragmented
- BERTopic: ~580 algorithmic topics : works are grouped meaningfully
- Provides a true data-driven analysis instead of manual labeling

HSDSLab
Human & Social Data Science

Topics

Topic Modelling - General Presentation

Data Separation

- Topic modelling conducted on all Q1 journals works for better coherence
- Extracting all BJP works by comparing common columns (title and others)
- Some works were lost during the process (26,115 → 22,308)

Topic Assignation

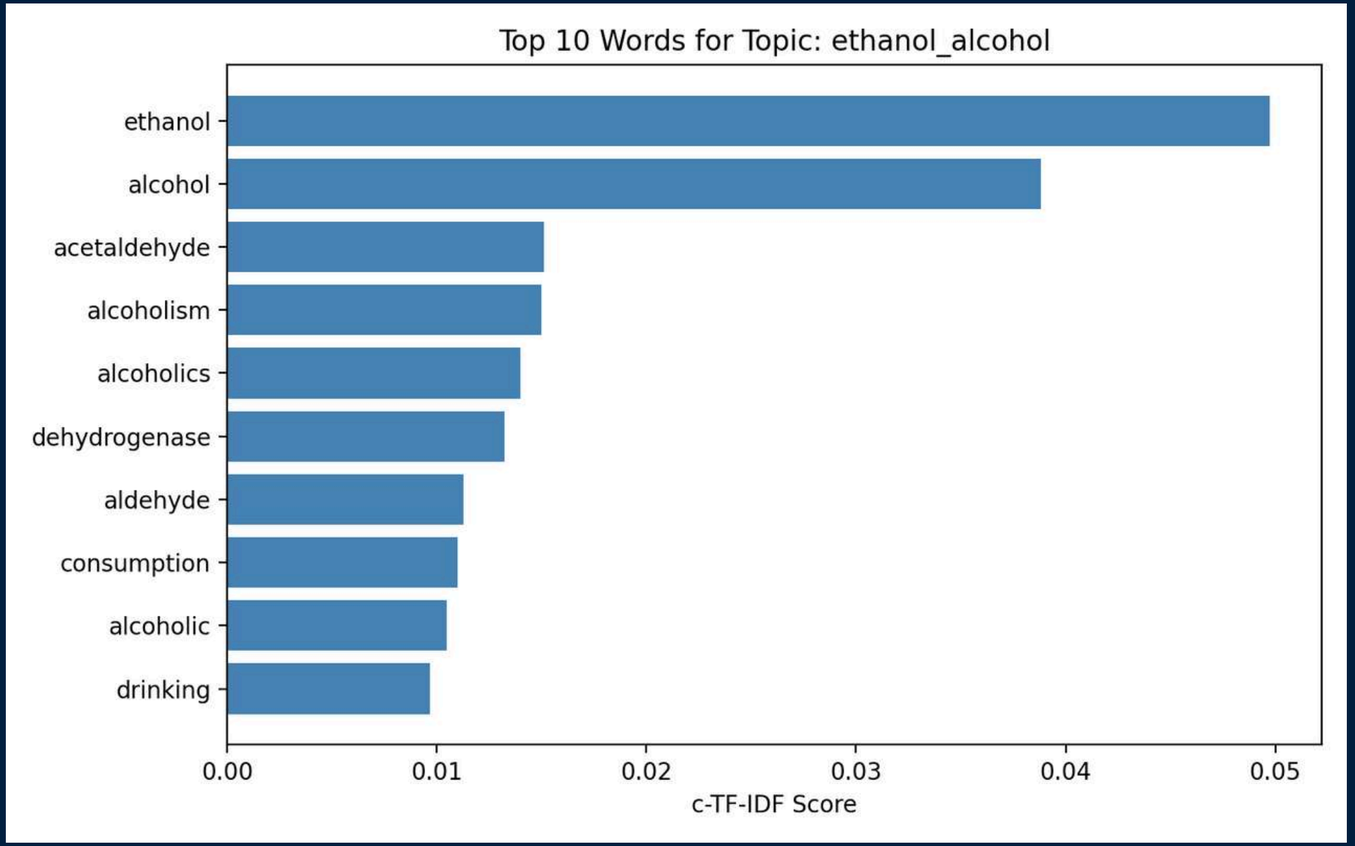
- 348 193 Q1 works without a topic assigned (689 391 → 341 198 works)
- 11052 BJP works without a topic assigned (22 308 → 11 256 works, 311 topics)

HSDSLab
Human & Social Data Science

Topics

Topic Modelling - Topic names

- The model creates topics and gives us the top words for each topic.
- Each word has a score showing how important it is.
- We simply take the top 2–3 words to create a clear topic label.



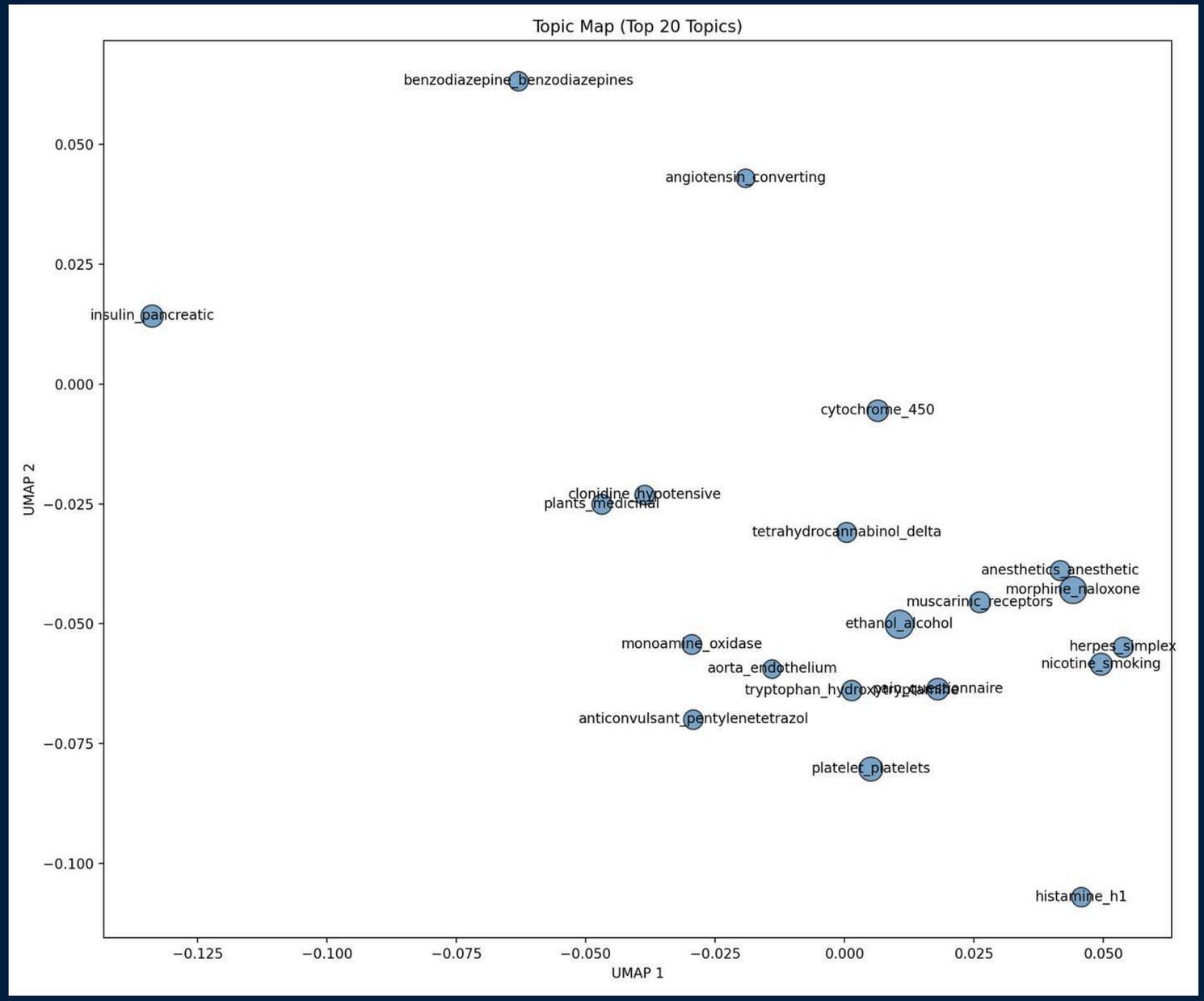
Exemple of words distribution for one topic

HSDSLab
Human & Social Data Science

Topics

Topic Modelling - Topics Similarity

- Top 20 topics by number of publications are shown for visualization purposes.
- Topics closer on the plot are more similar in content.

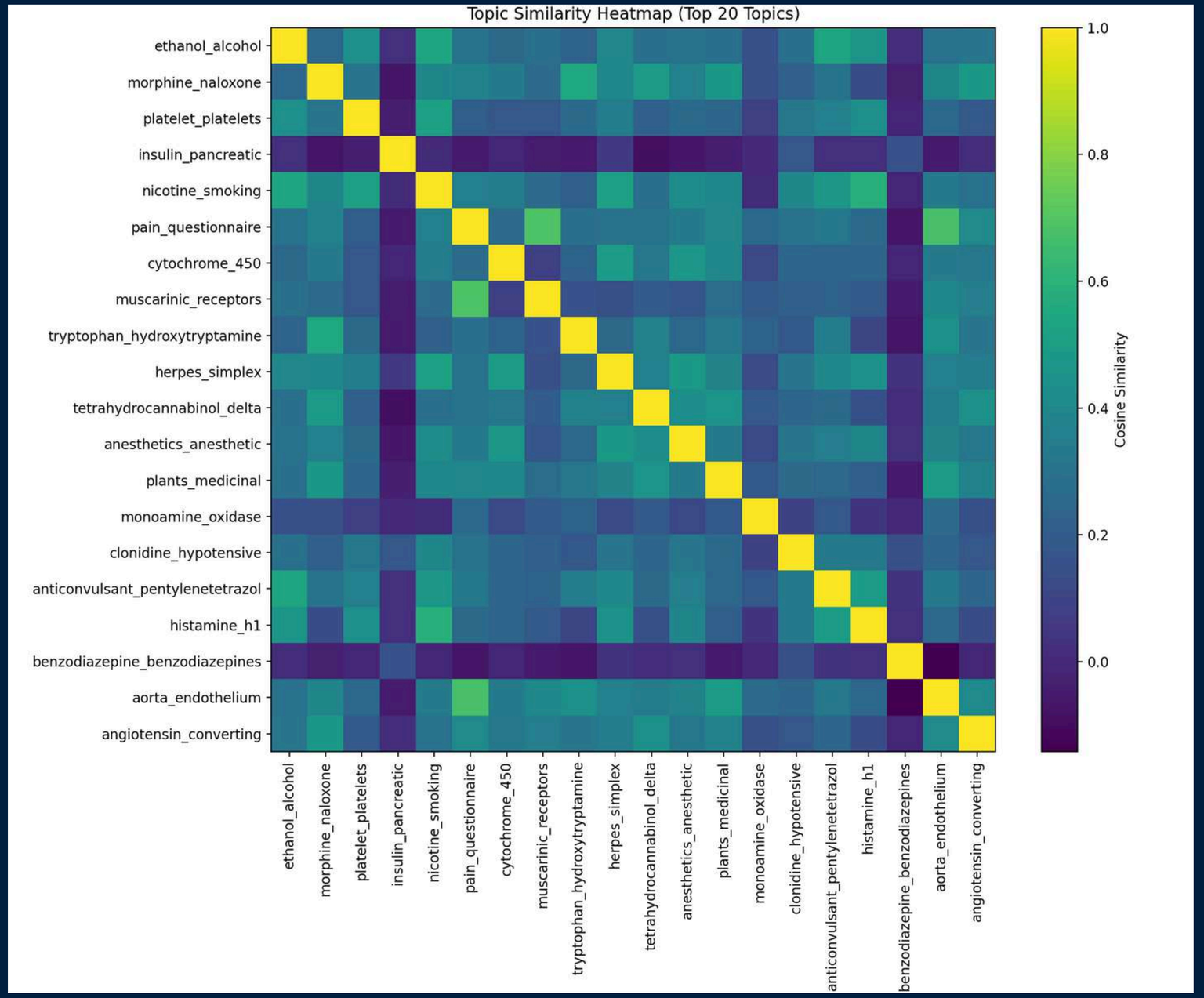


HSDSLab
Human & Social Data Science

Topics

Topic Modelling - Topics Similarity

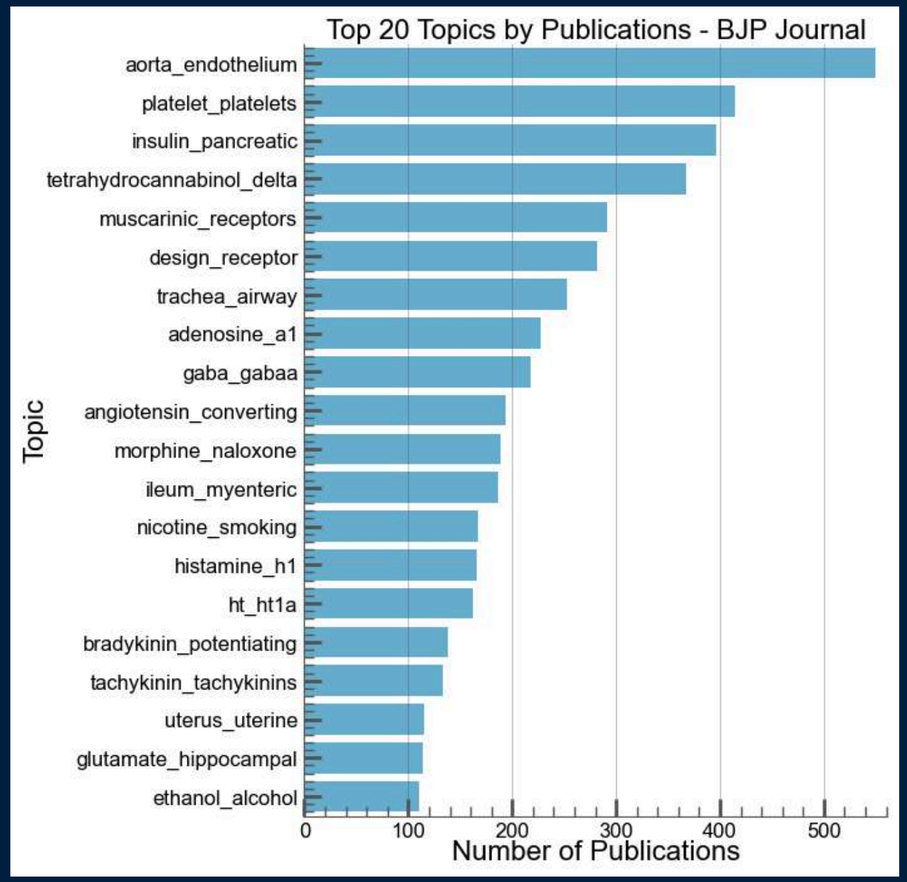
- Top 20 topics by number of publications are shown for visualization purposes.

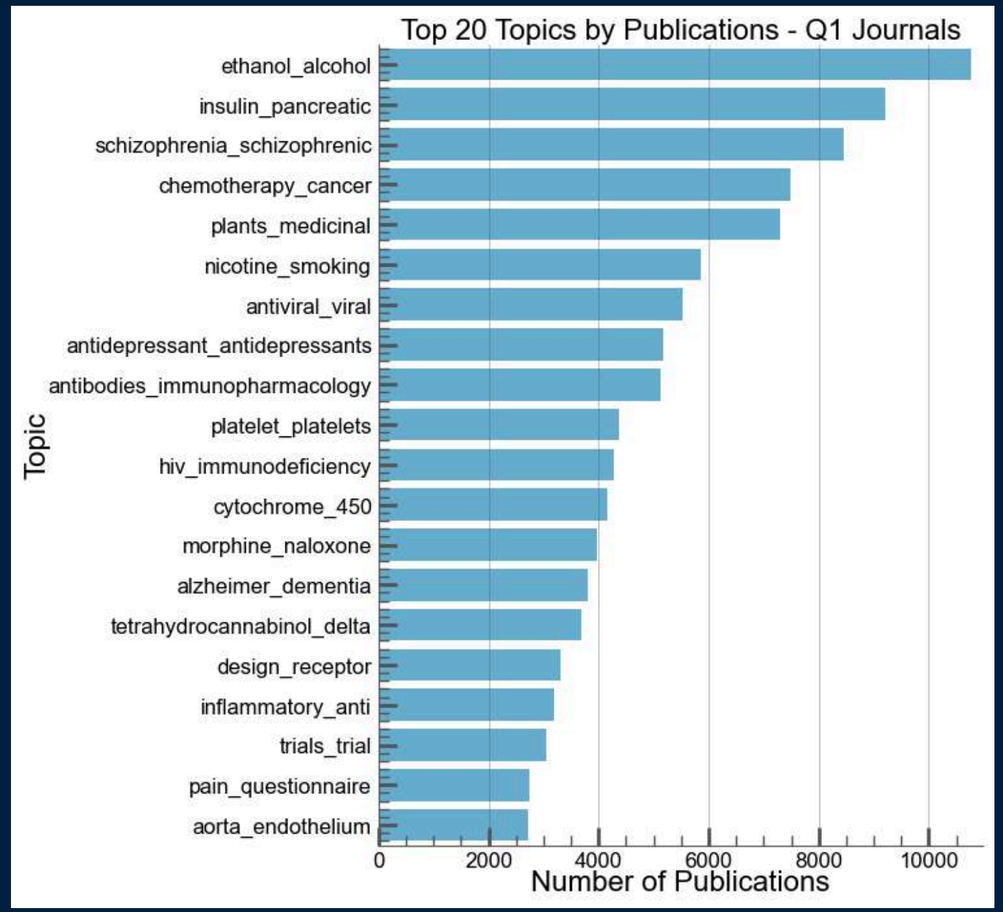


HSDSLab
Human & Social Data Science

Topics

Topic Modelling -
Top Publications (Cumulative)

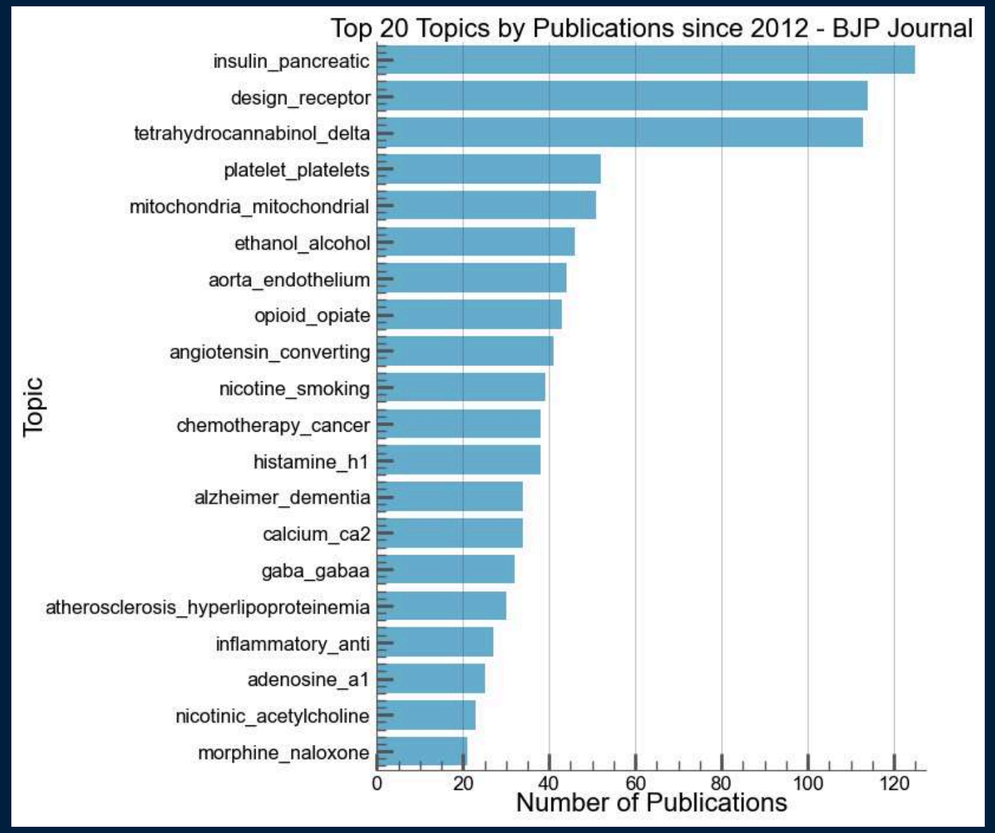


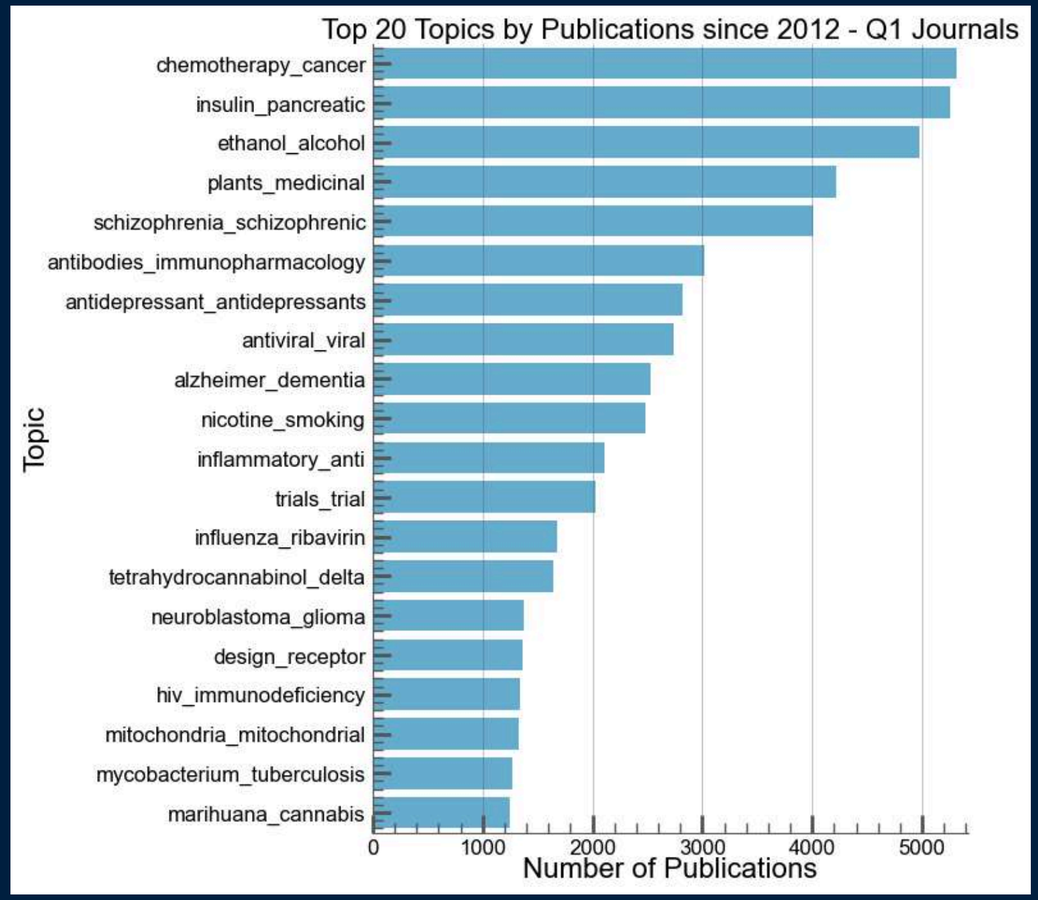


HSDSLab
Human & Social Data Science

Topics

Topic Modelling -
Top Publications (since 2012)

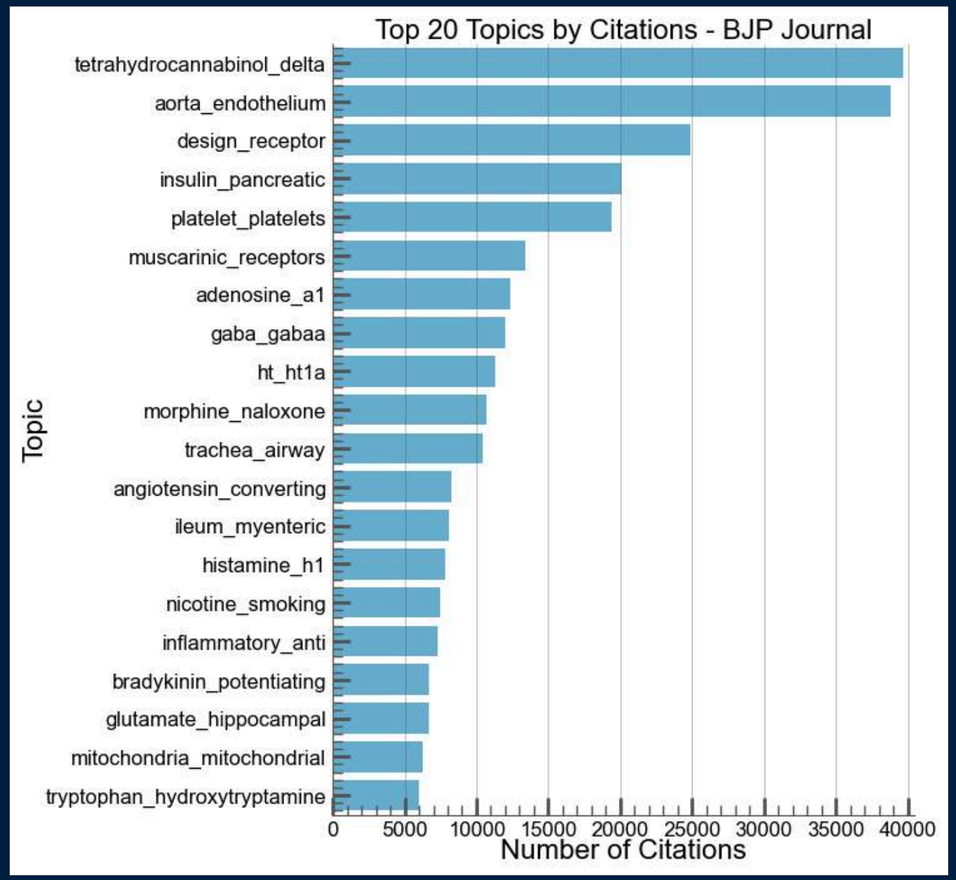


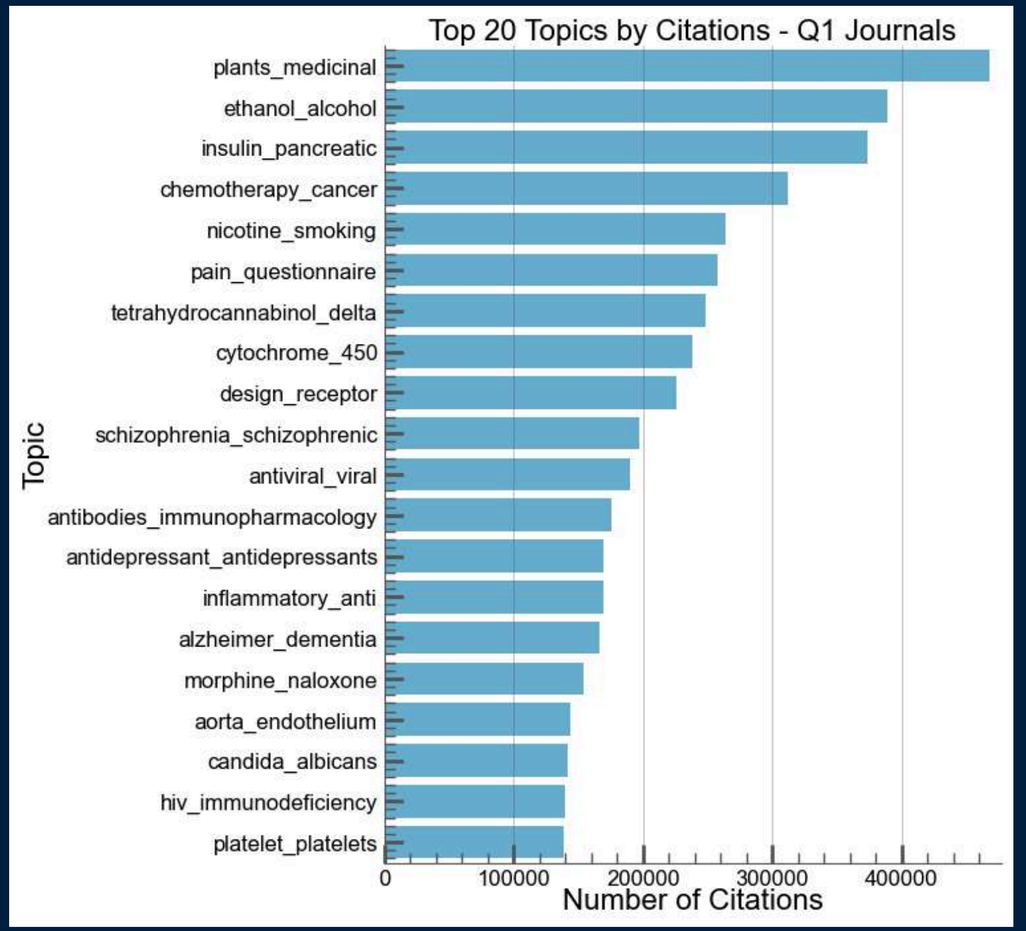


HSDSLab
Human & Social Data Science

Topics

Topic Modelling -
Top Citations (Cumulative)

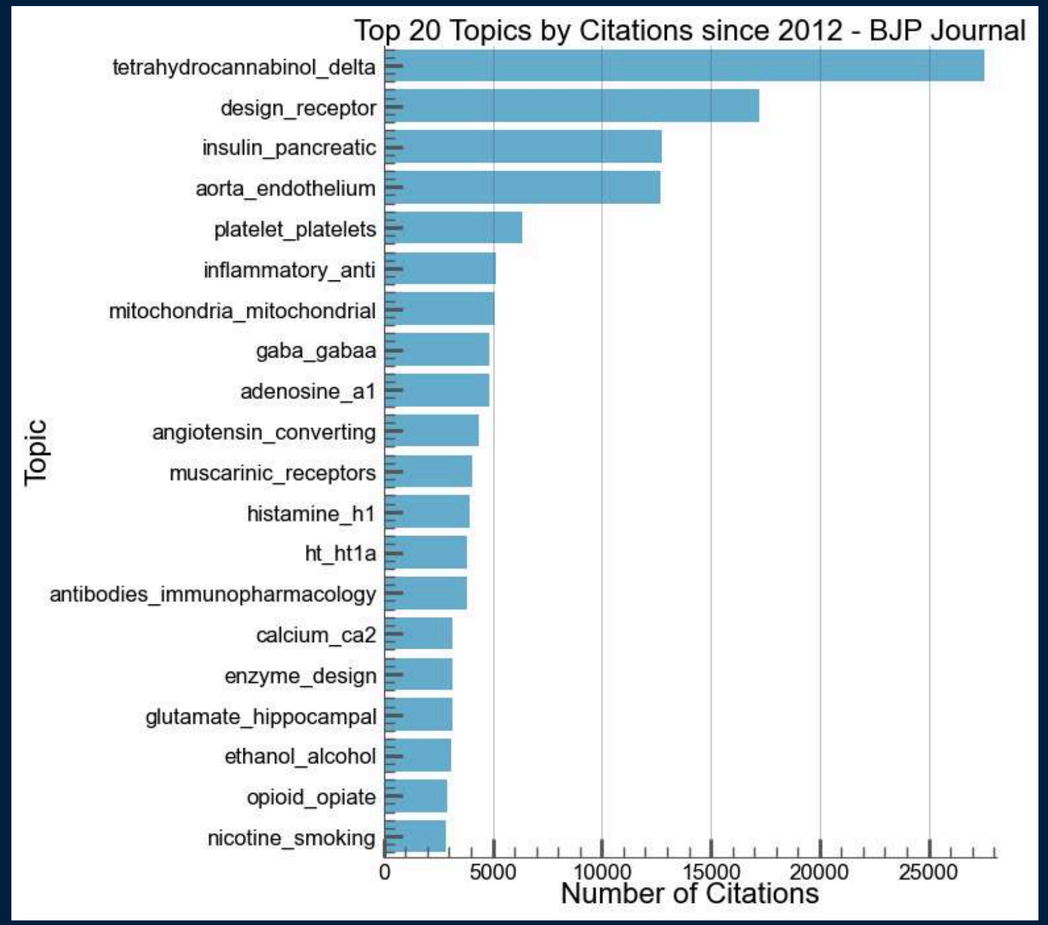


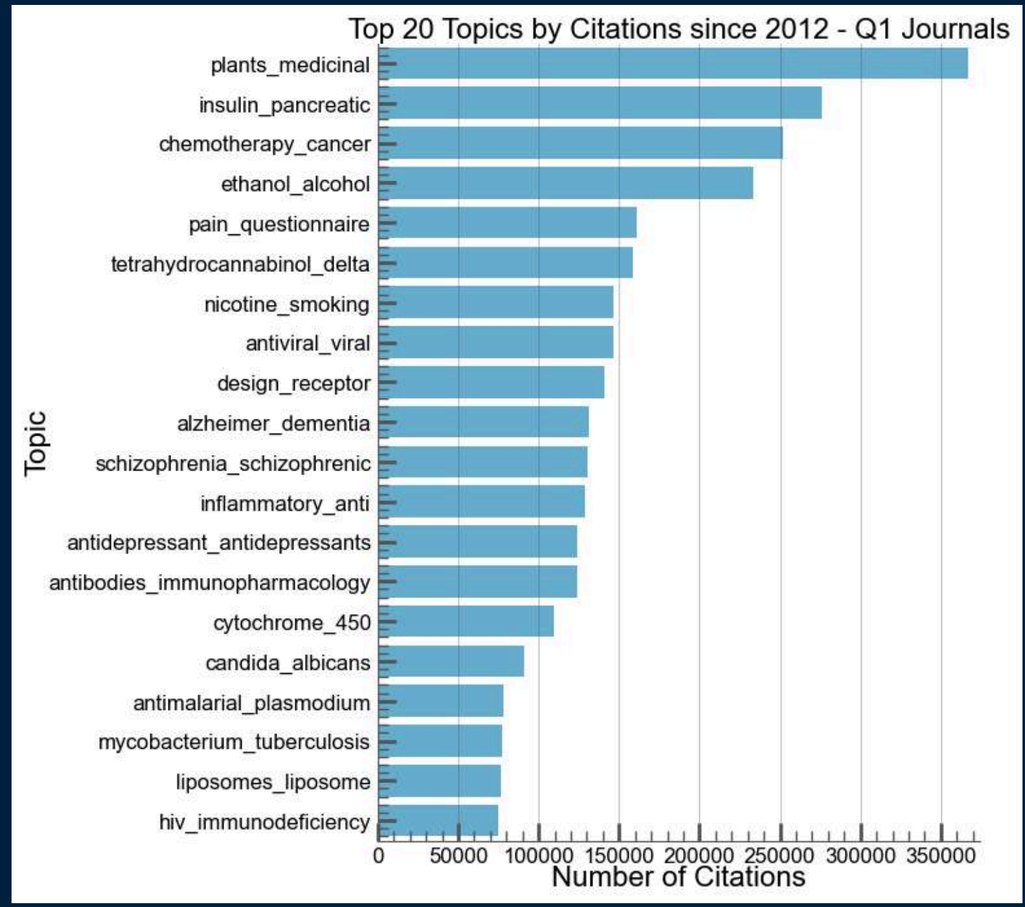


HSDSLab
Human & Social Data Science

Topics

Topic Modelling -
Top Citations (since 2012)



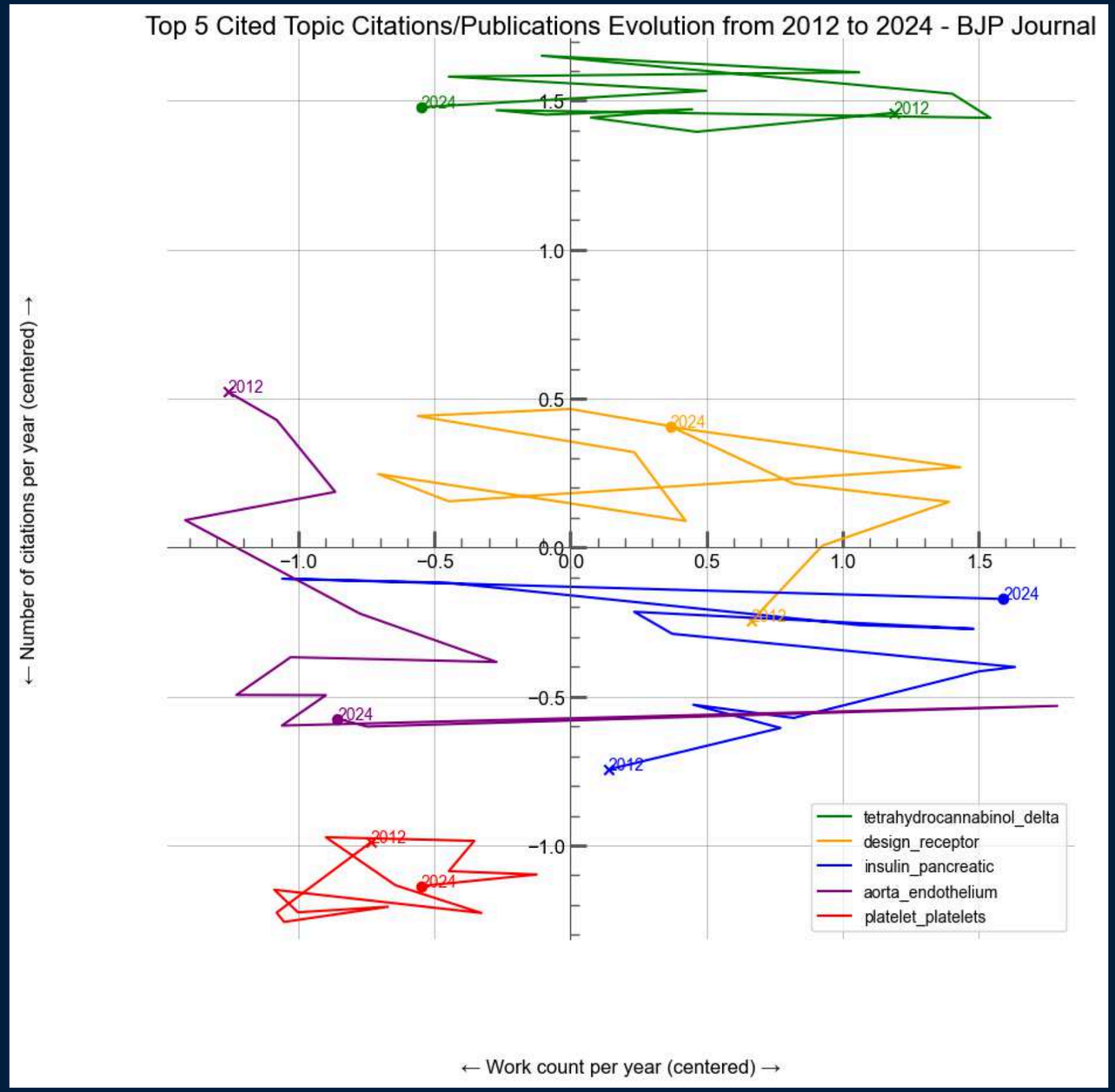


HSDSLab
Human & Social Data Science

Topics

Topic Modelling - Top 5 cited BJP topics since 2012 Citations/Publications

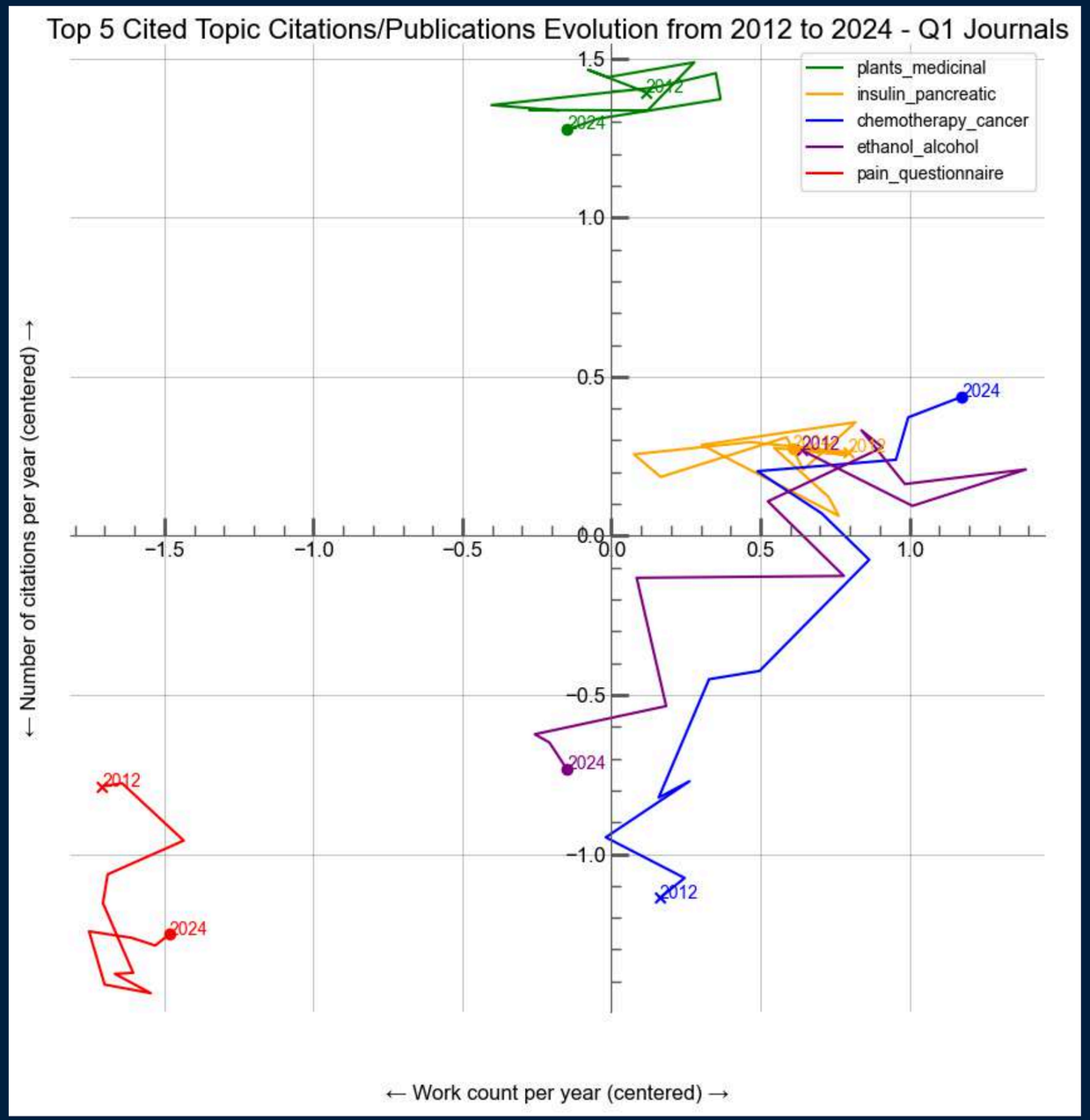
- tetrahydrocannabinol remains at the top
- Small increase for insulin_pancreatic
- Small decline for aorta_endothelium



Topics

Topic Modelling - Top 5 Q1 cited topics since 2012 Citations/Publications

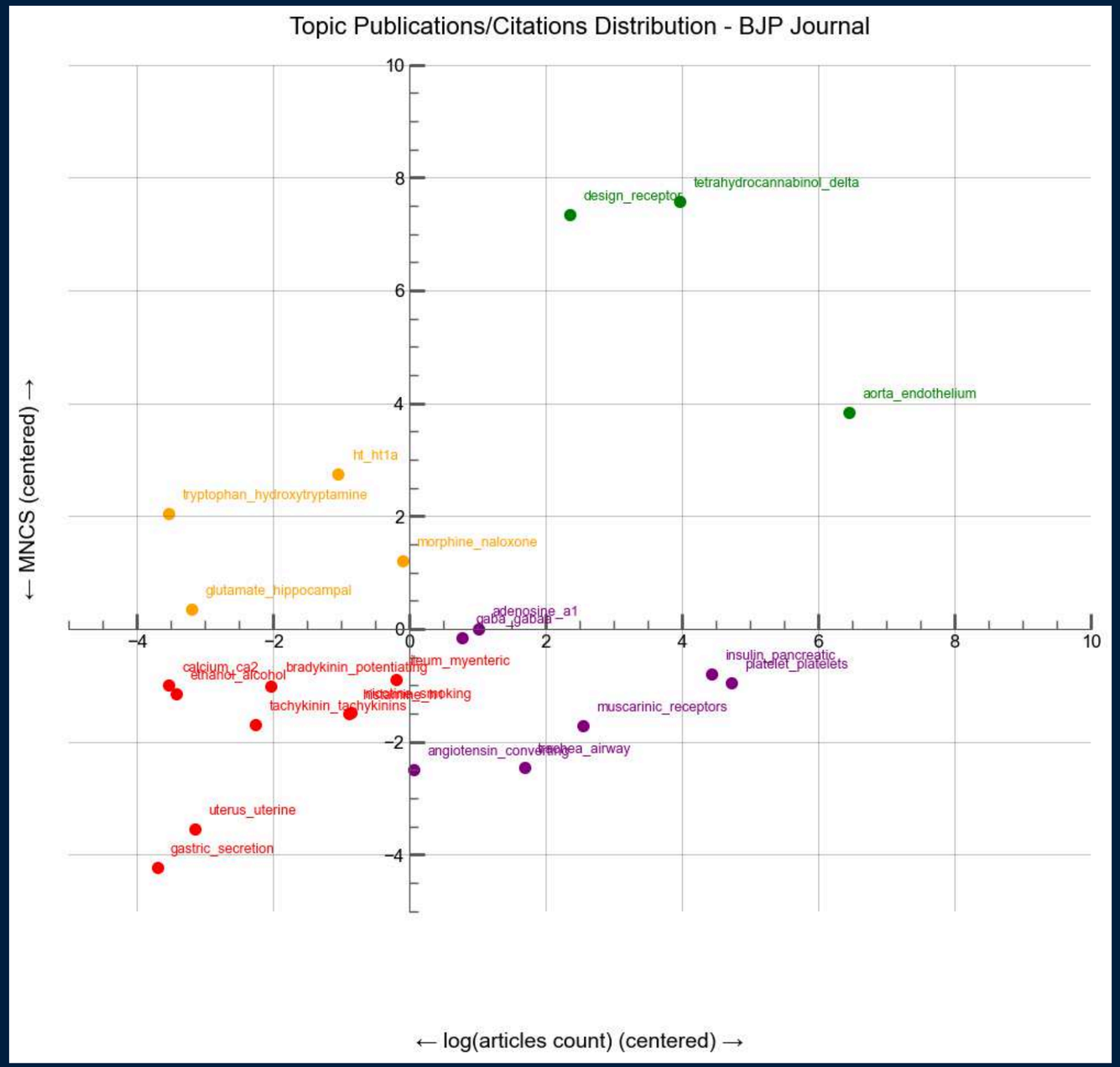
- plants_medicinal remains at the top
- Small increase for chemotherapy_cancer
- Small decline for ethanol_alcool



Topics

Topic Modelling - General topics BJP MNCS/Publications Distribution

- Topics with 100+ works are shown for visualization purposes.
- “Over-Published” topics : (platelets, muscarinic_receptors, ...)
- “Under-Published” Topics : (ht_ht1a, hydroxytryptamine, ...)

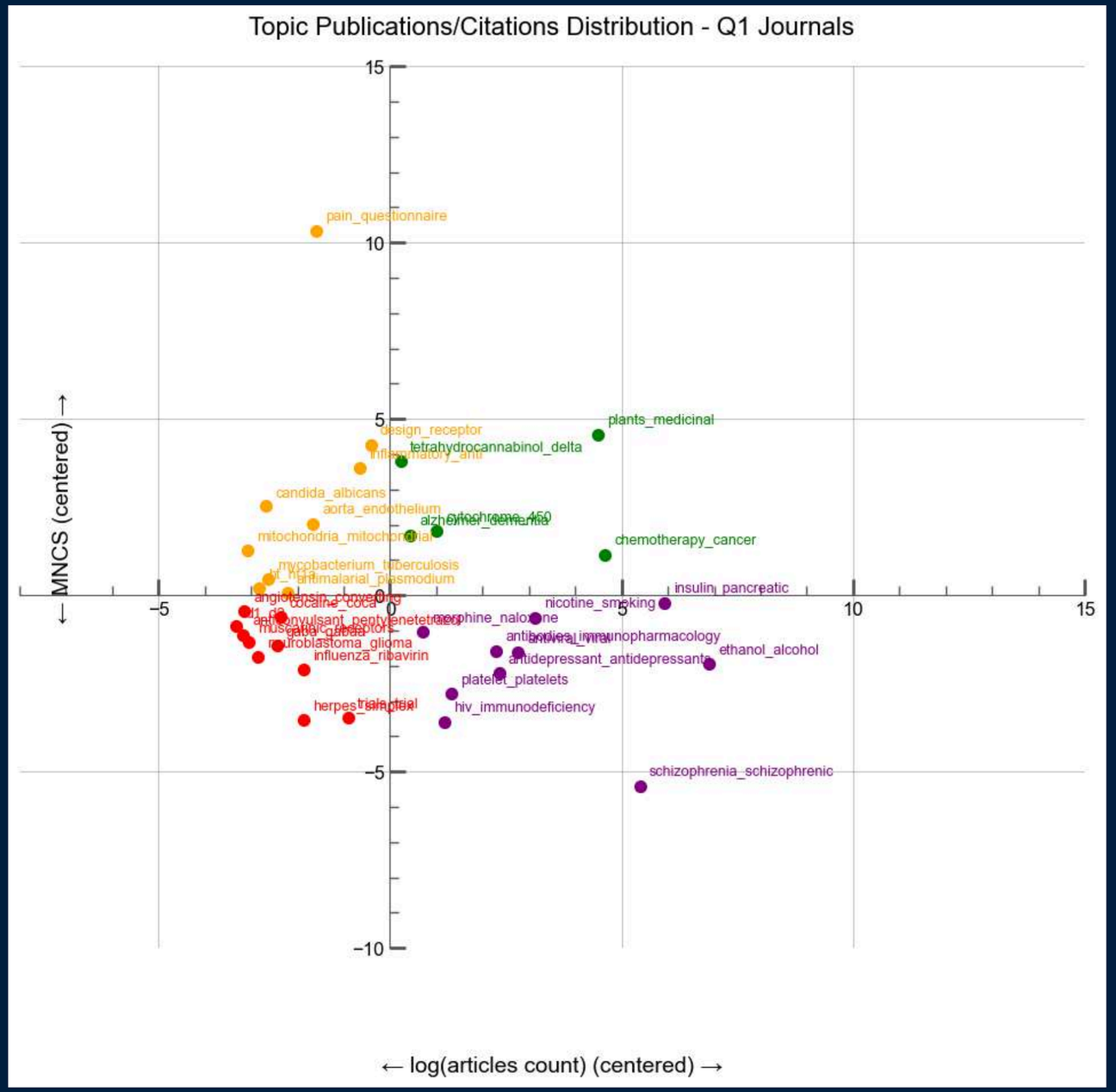


HSDSLab
Human & Social Data Science

Topics

Topic Modelling - General topics Q1 MNCS/Publications Distribution

- Topics with 2000+ works are shown for visualization purposes.
- “Over-Published” topics : (schizophrenia, ethanol_alcohol, ...)
- “Under-Published” Topics : (design_receptor, pain_questionnaire, ...)



HSDSLab
Human & Social Data Science

Modelling

Modelling

Modelling - General Presentation

Why Modelling ?

- Publications differ in many ways (authors, institutions, topics)
- All these differences can affect how often a paper is cited
- It’s hard to understand this just by looking at the data
- Modelling helps us see which factors matter most

How does it work ?

- We use machine-learning models trained on the characteristics of each publication
- A machine-learning model is a tool that learns patterns from examples and links them to an outcome like MNCS
- We measure how strongly each factor contributes to the predicted MNCS

HSDSLab
Human & Social Data Science

Modelling

Modelling by Decade - Presentation

Model Formula :

$$MNCS = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon$$

Modeling Approach

- We divide the data by decade and fit a simple **linear regression** model for each period
- This approach helps us capture trends and “hypes” in different decades, especially for topics.

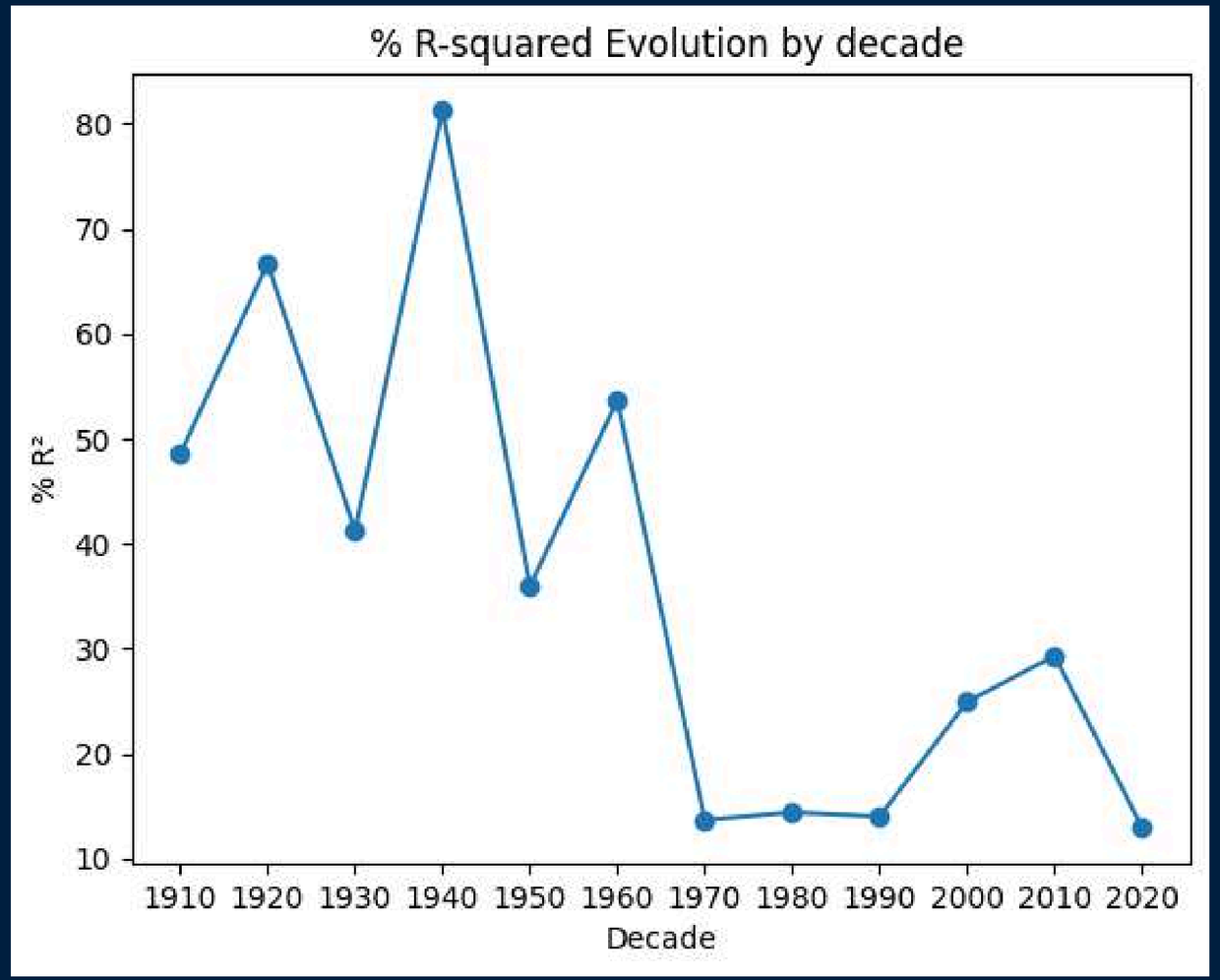
- $X_1, X_2, \ldots, X_n$ are features of the publications (authors, institutions, topics, etc...)
- $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients estimated by the model
- ϵ is the error term (not taken into consideration)

HSDSLab
Human & Social Data Science

Modelling

Modelling by Decade - R^2 metric

- % R^2 shows how much MNCS variation the model explains per decade
- High early on due to few publications
- Lower in 1970s–1990s & 2020s (~13–15%), still capturing meaningful patterns
- Higher in 2000s–2010s (~25–30%), reflecting stronger trends and topic “hypes”



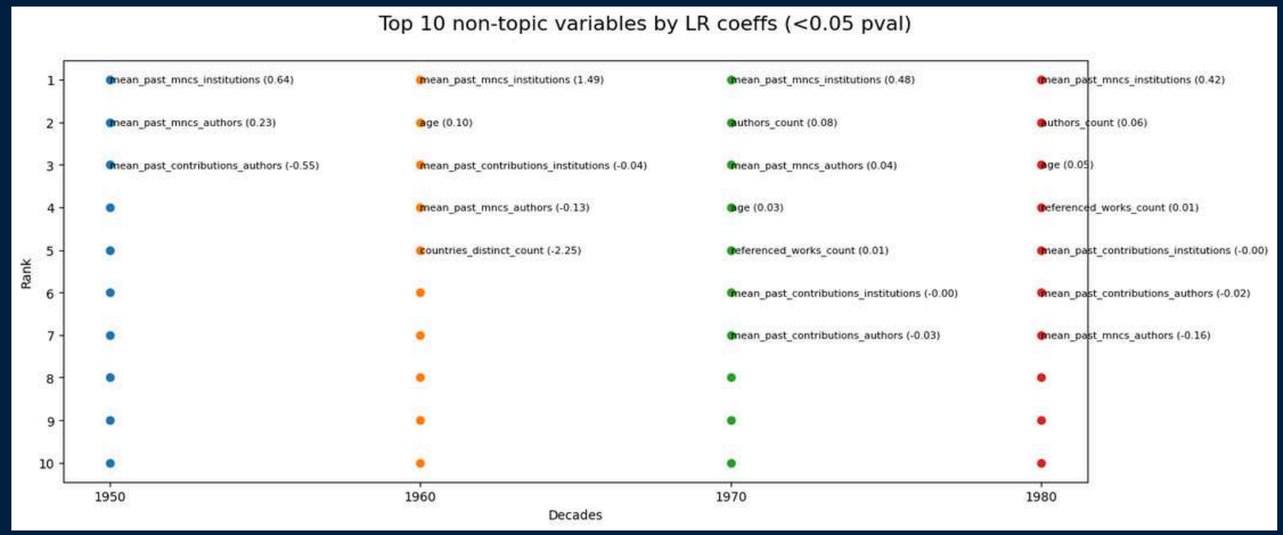
HSDSLab
Human & Social Data Science

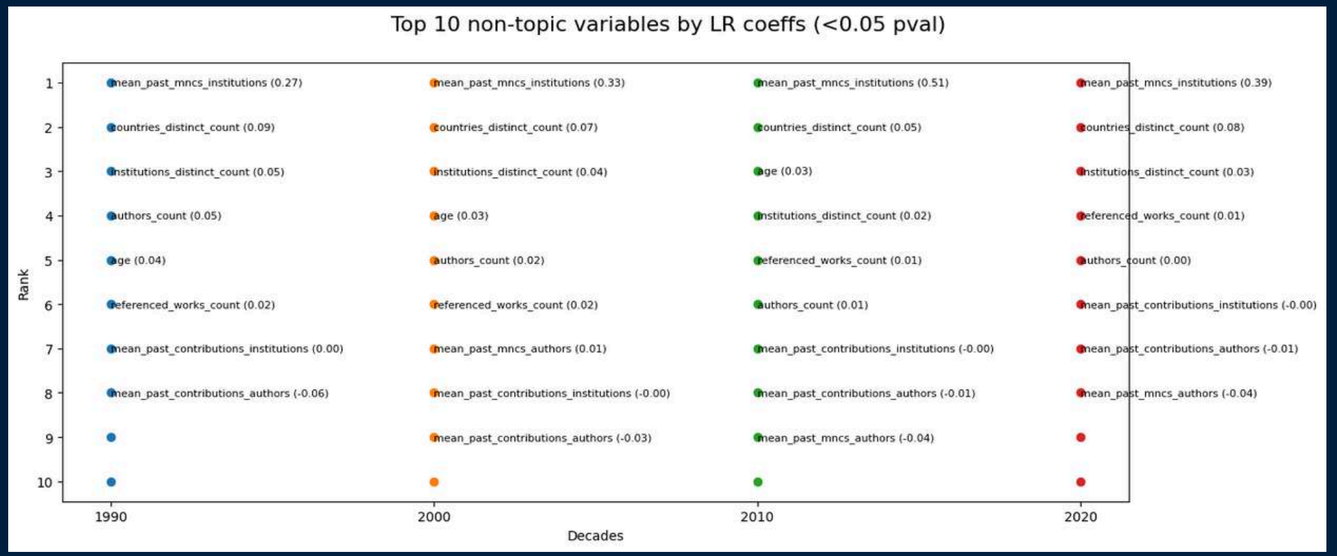
Modelling

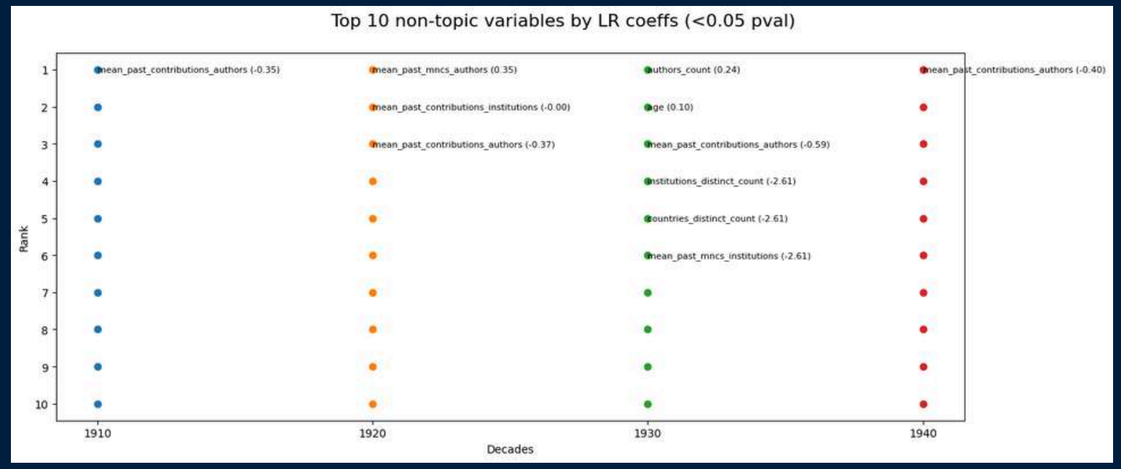
Modelling by Decade - Other Variables

Other = non-topic
< 0.05 pval : Signicative Value

Mean past MNCS Institutions seems to be the most important variable (Higher coeff)





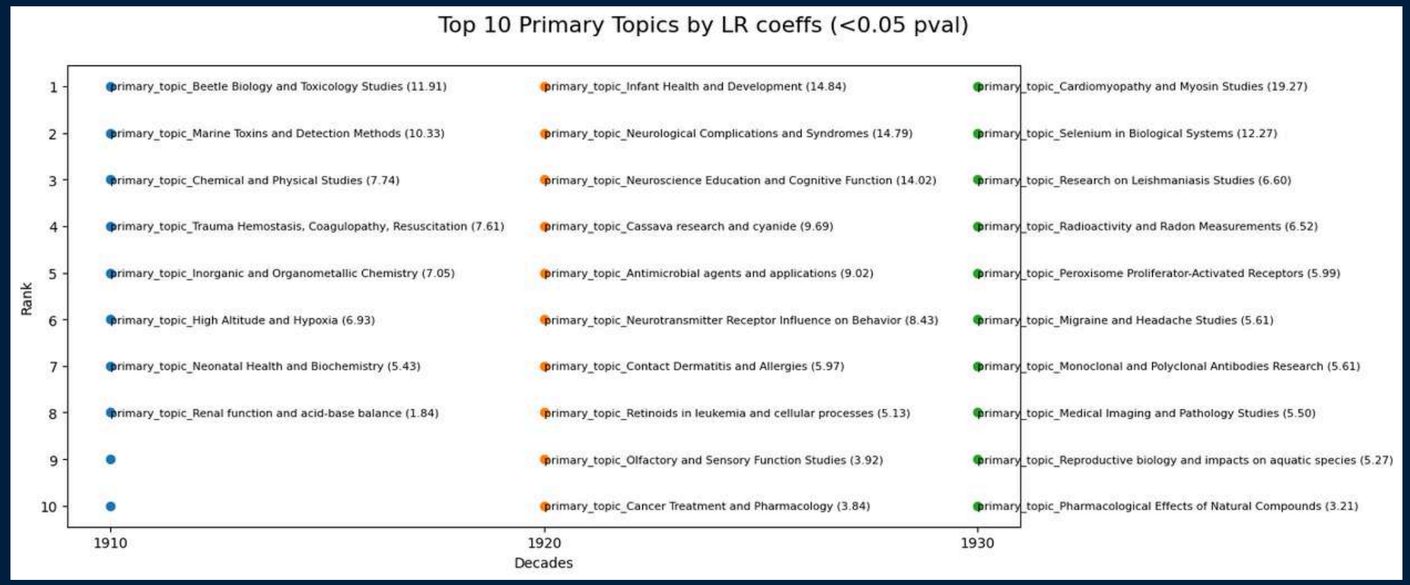


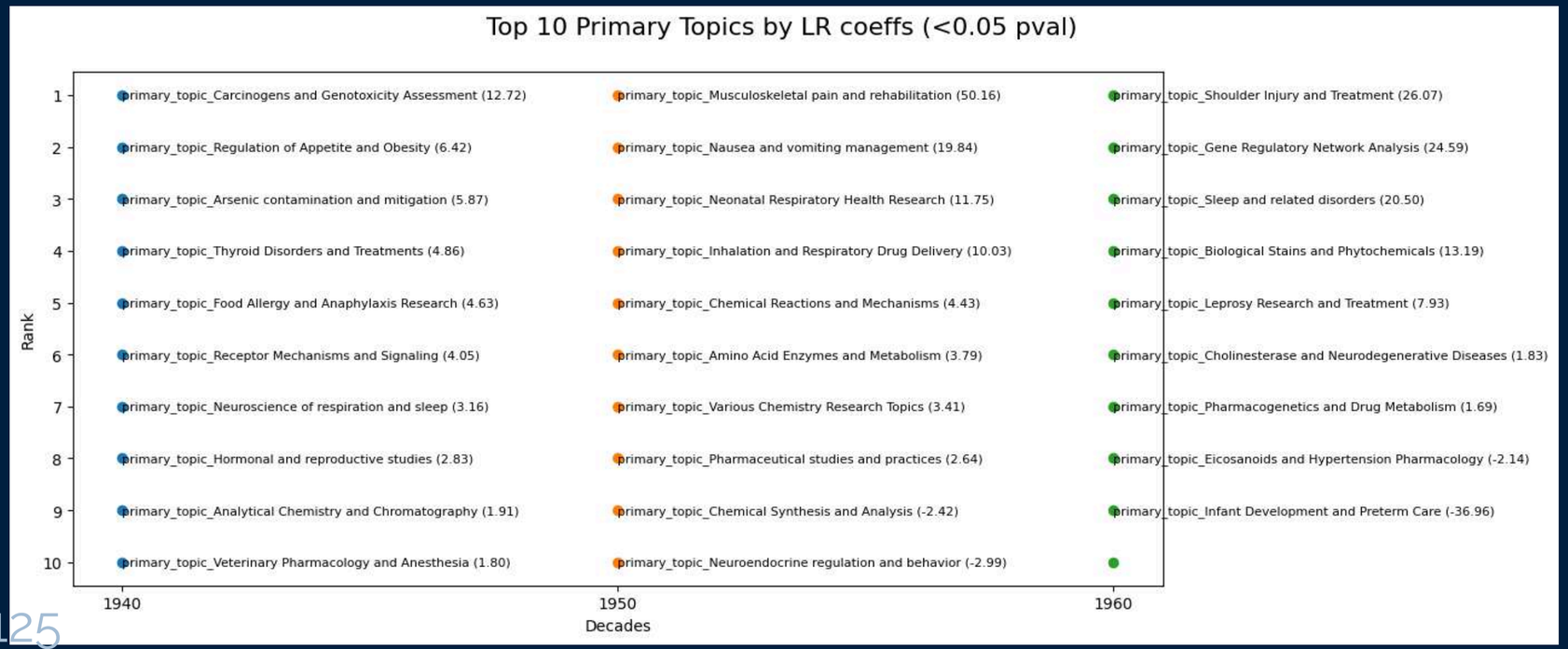
HSDSLab
Human & Social Data Science

Modelling

Modelling by Decade - Primary Topics

< 0.05 pval : Signicative Value



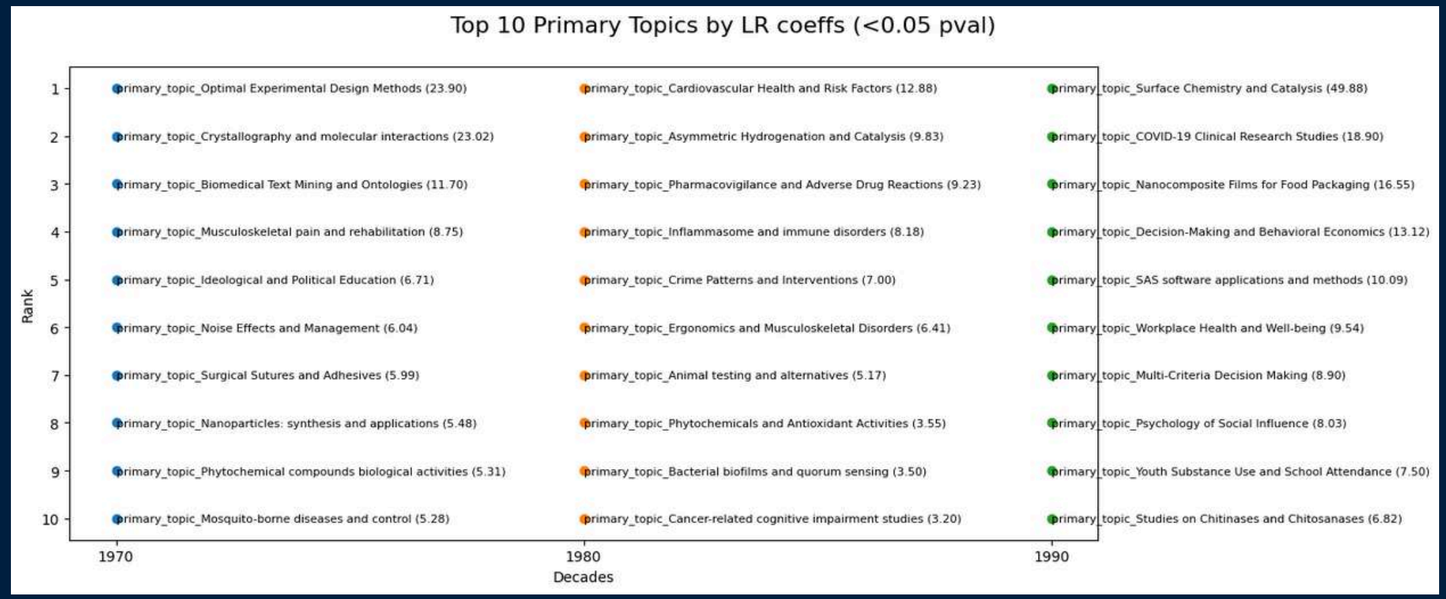


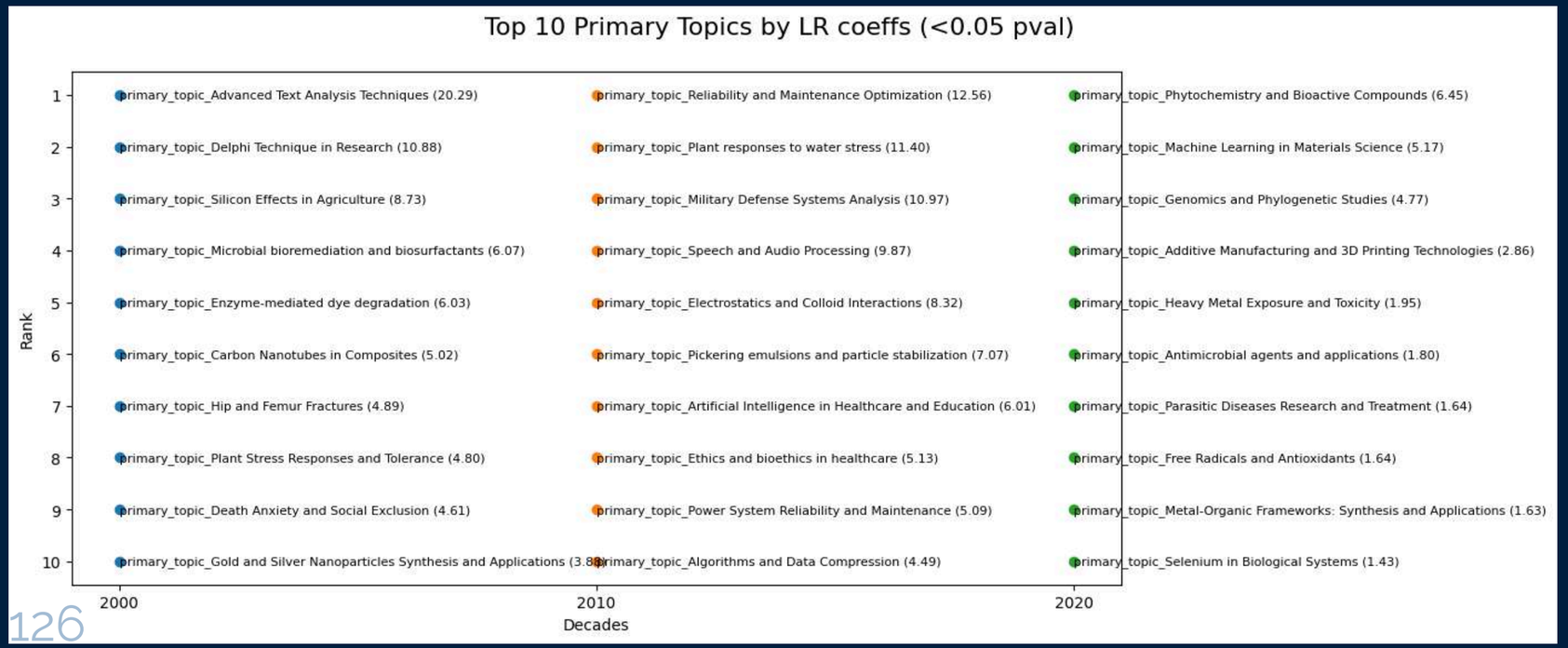
HSDSLab
Human & Social Data Science

Modelling

Modelling by Decade - Primary Topics

< 0.05 pval : Signicative Value



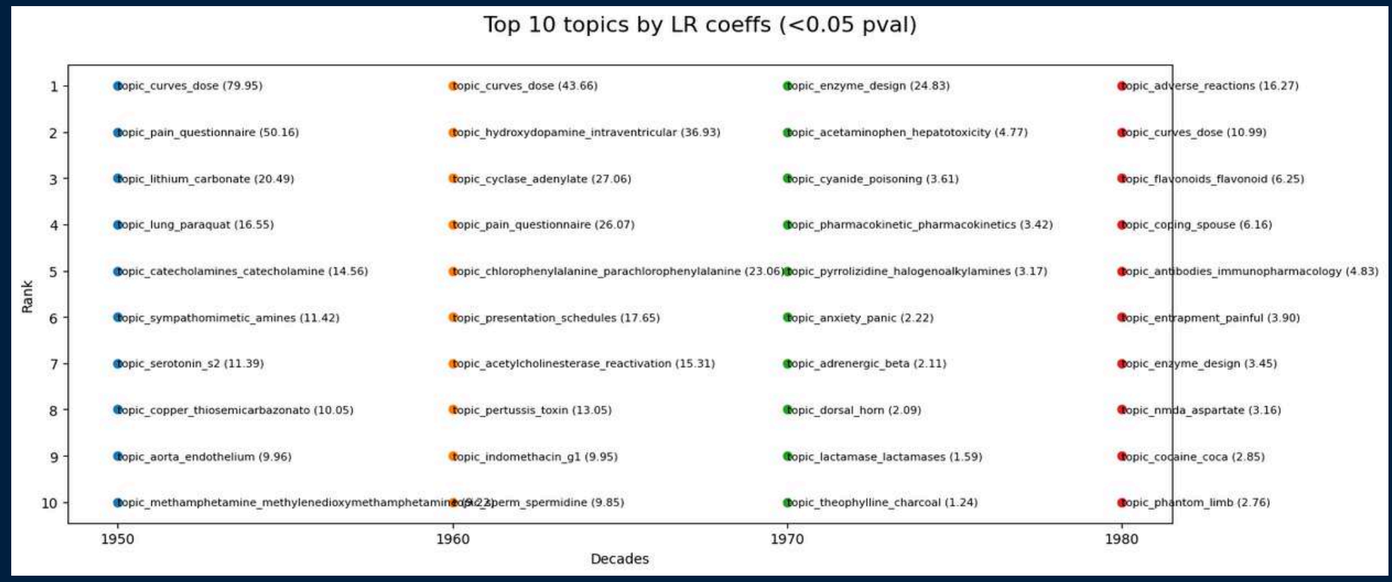


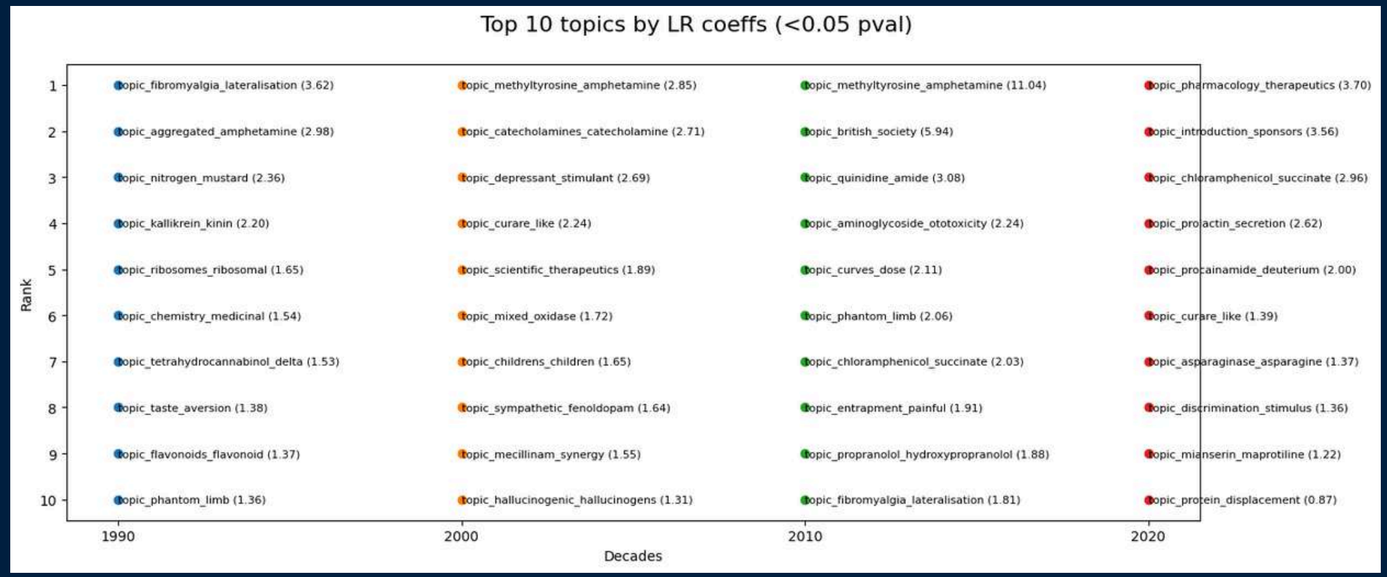
HSDSLab
Human & Social Data Science

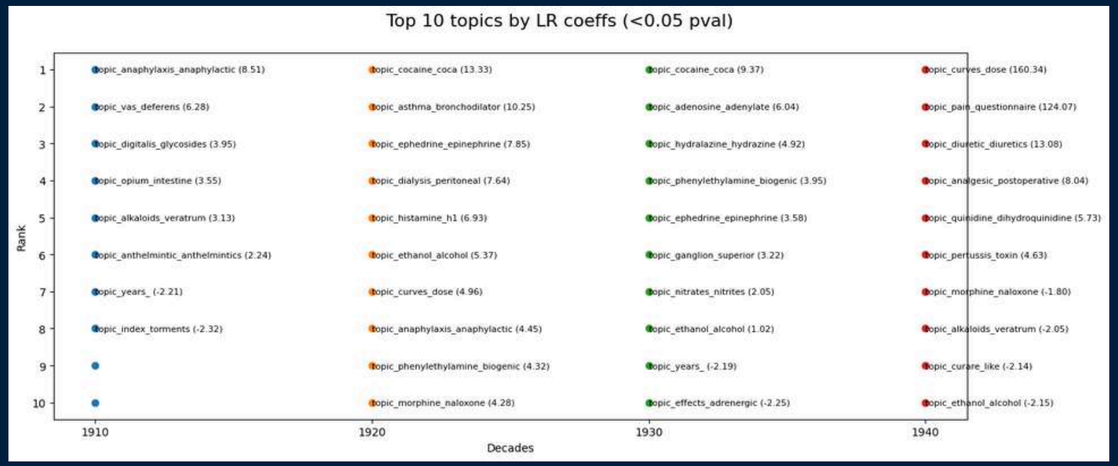
Modelling

Modelling by Decade - Modelling Topics

< 0.05 pval : Signicative Value





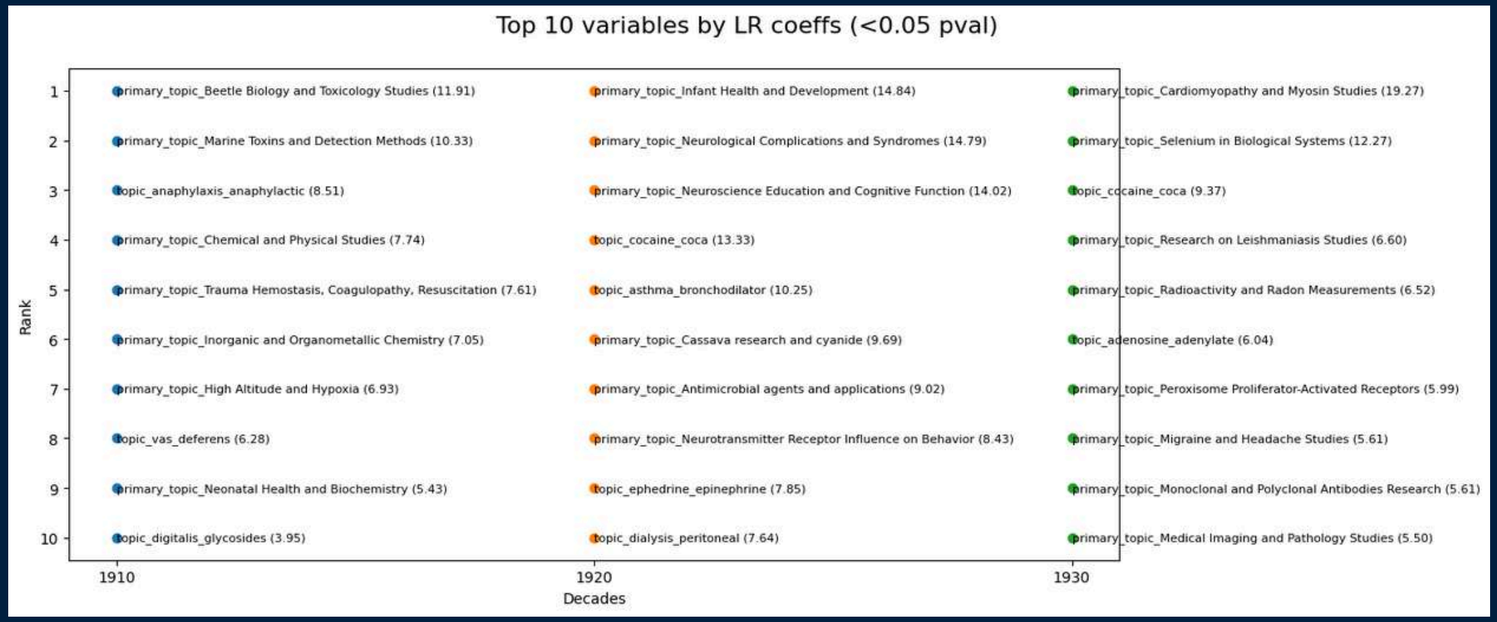


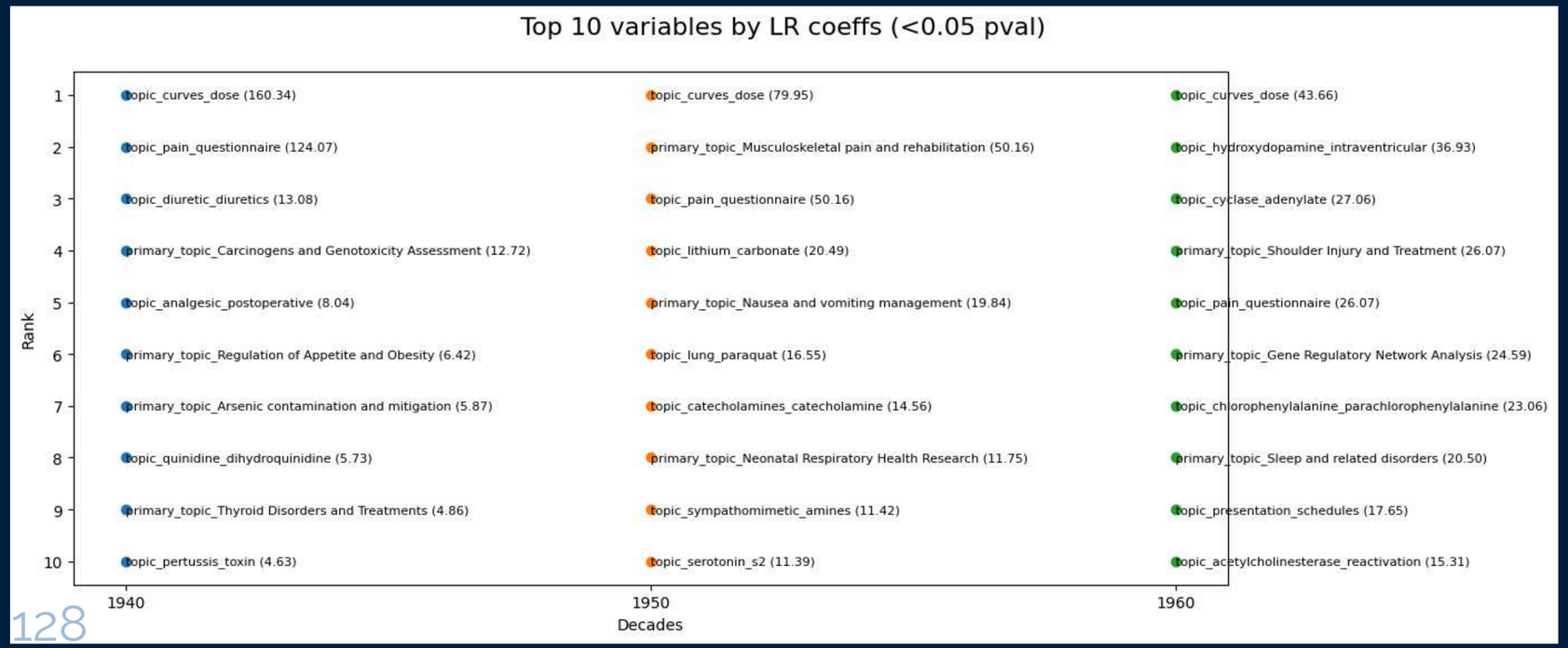
HSDSLab
Human & Social Data Science

Modelling

Modelling by Decade - All Variables

< 0.05 pval : Signicative Value



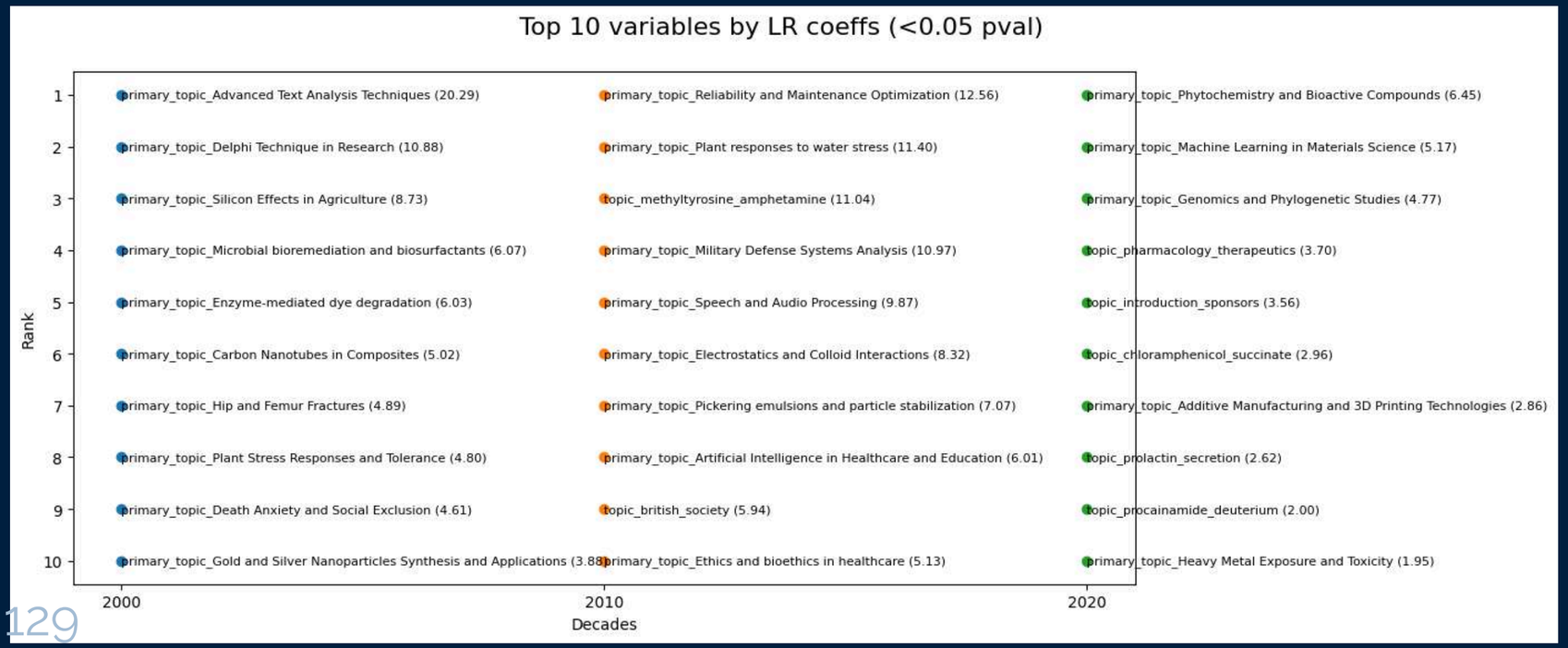


HSDSLab
Human & Social Data Science

Modelling

Modelling by Decade - All Variables

< 0.05 pval : Signicative Value



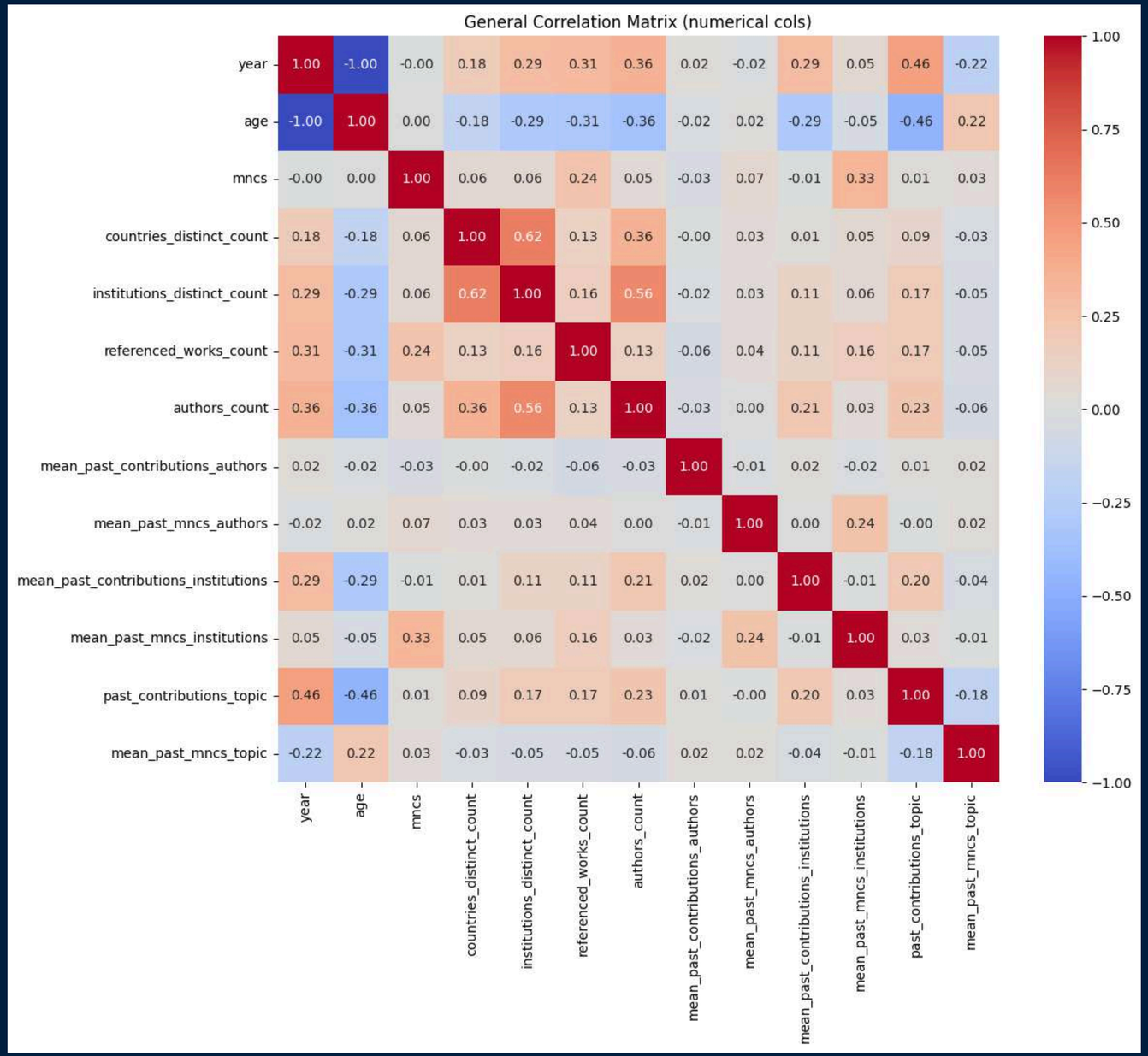
HSDSLab
Human & Social Data Science

Modelling

General Modelling - Presentation

- We apply several machine-learning models to the full dataset
- Focus is on variables beyond specific topics (authors, institutions, collaborations, past performance)
- We compare the models' performance and the importance of their features

Correlation Matrix : Values show how strongly two variables are related and in which direction (ex: positive relation between mncs/references-mncs past Institutions)



HSDSLab
Human & Social Data Science

Modelling

General Modelling - Presentation

Methodology

- ML models predict the MNCS quantile category of each work : 0 - 0.25 = **0**; 0.25 - 0.5 = **1** 0.5 - 0.7 = **2**; 0.7 - 0.9 = **3**; 0.9 - 1 = **4**)

- Training the models on 550 000+ works, testing (predictions) on 138 000+ works

Performance Metrics

- Recall – fraction of true samples correctly found (confusion matrix)

- Precision – fraction of correct predictions per class (scatter plots)

- Accuracy – overall correct predictions

- RMSE – average distance between predicted & true classes

- QWK – measures ordinal agreement; 1 = perfect, 0 = random, >0 = better than chance

HSDSLab
Human & Social Data Science

Modelling

General Modelling - Logistic Regression

Like linear regression, but outputs probabilities between 0 and 1

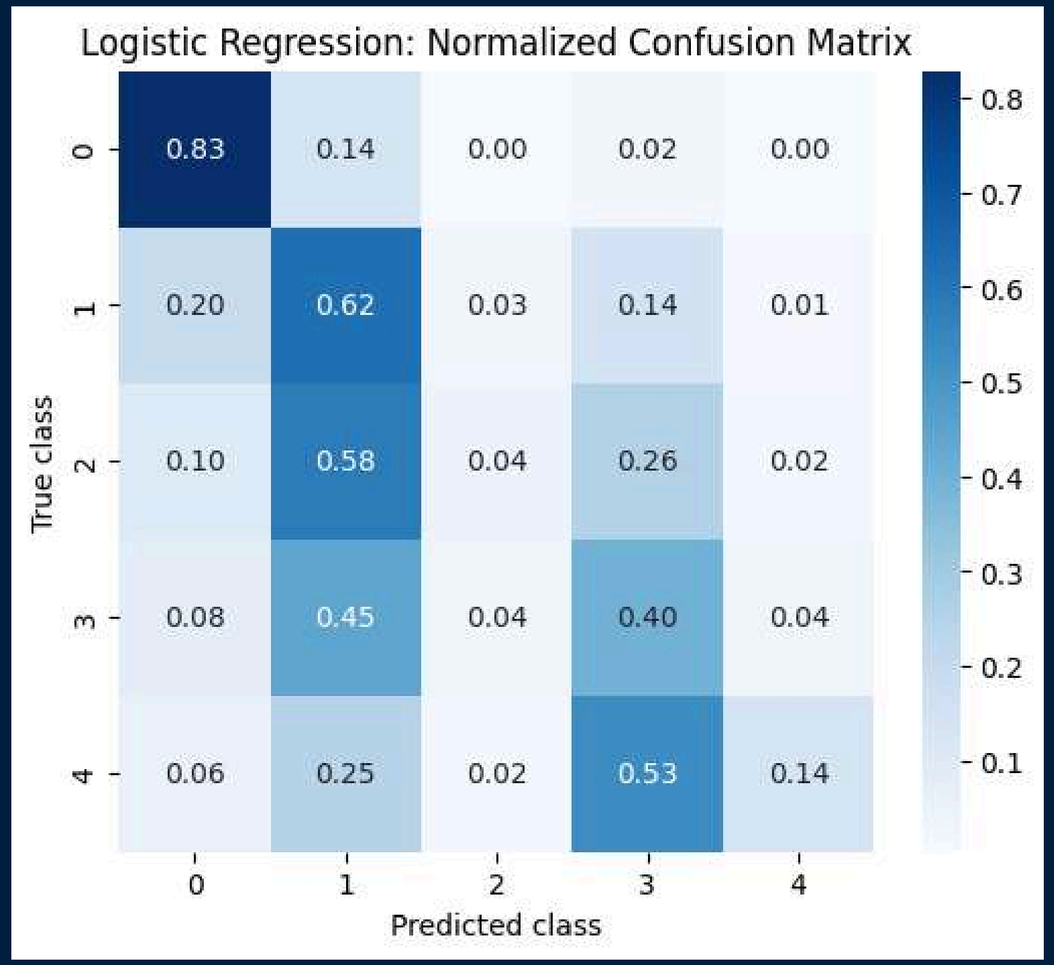
```
Classification report:
              precision    recall  f1-score   support

           0       0.69      0.83      0.75     34481
           1       0.37      0.62      0.46     34489
           2       0.31      0.04      0.07     27613
           3       0.35      0.40      0.37     27507
           4       0.50      0.14      0.22     13783

    accuracy                           0.46    137873
   macro avg       0.44      0.41      0.38    137873
weighted avg       0.45      0.46      0.42    137873

RMSE: 1.2207
QWK (Quadratic Weighted Kappa): 0.5471
```

Confusion Matrix (recall visualization) :



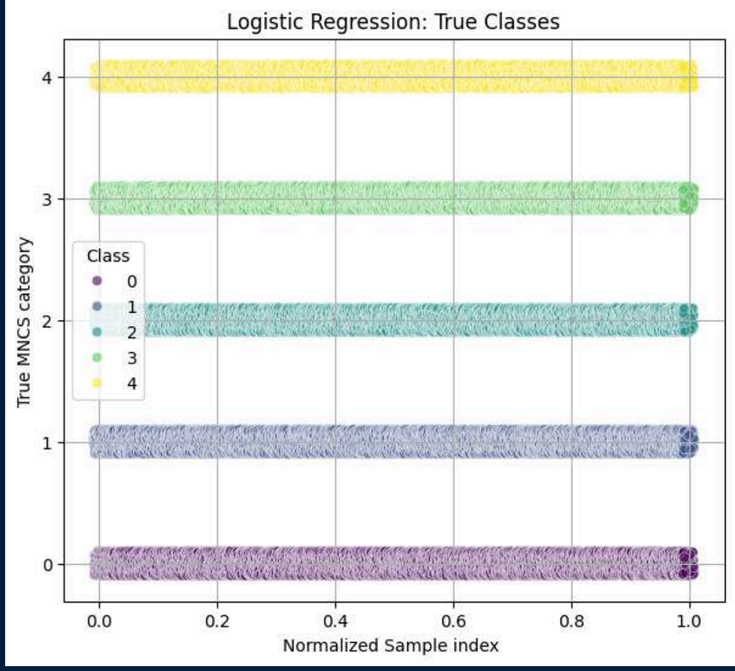
HSDSLab
Human & Social Data Science

Modelling

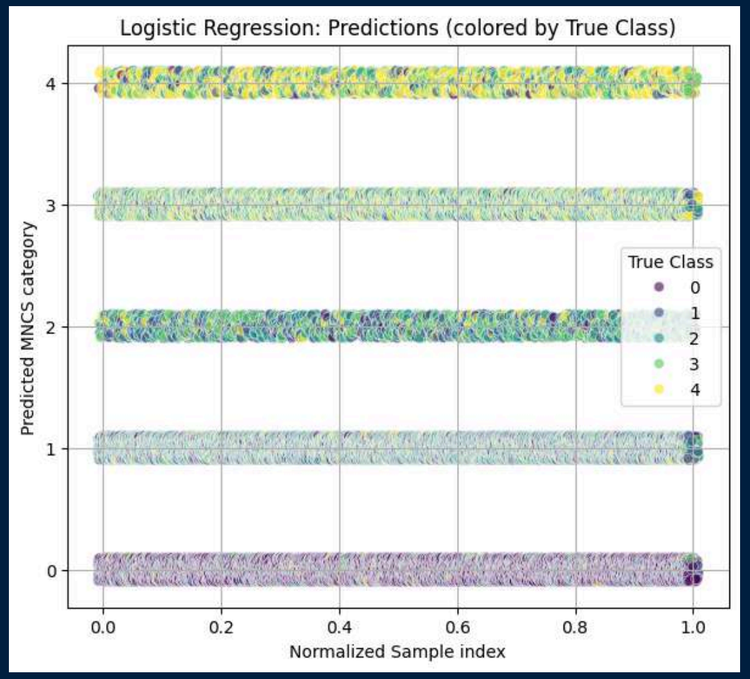
General Modelling - Logistic Regression

Like linear regression, but outputs probabilities between 0 and 1

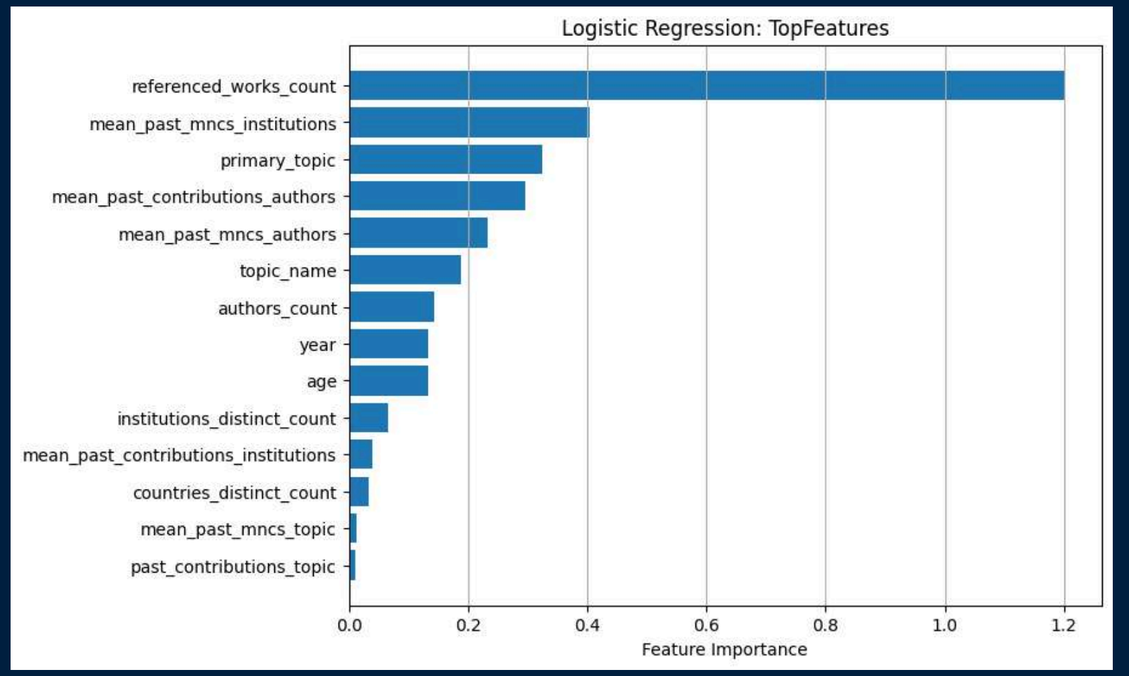
Reference ;



True classes vs Predictions (precision) :



Top Features :



HSDSLab
Human & Social Data Science

Modelling

General Modelling - Random Forest

A collection of decision trees whose outputs are averaged or voted

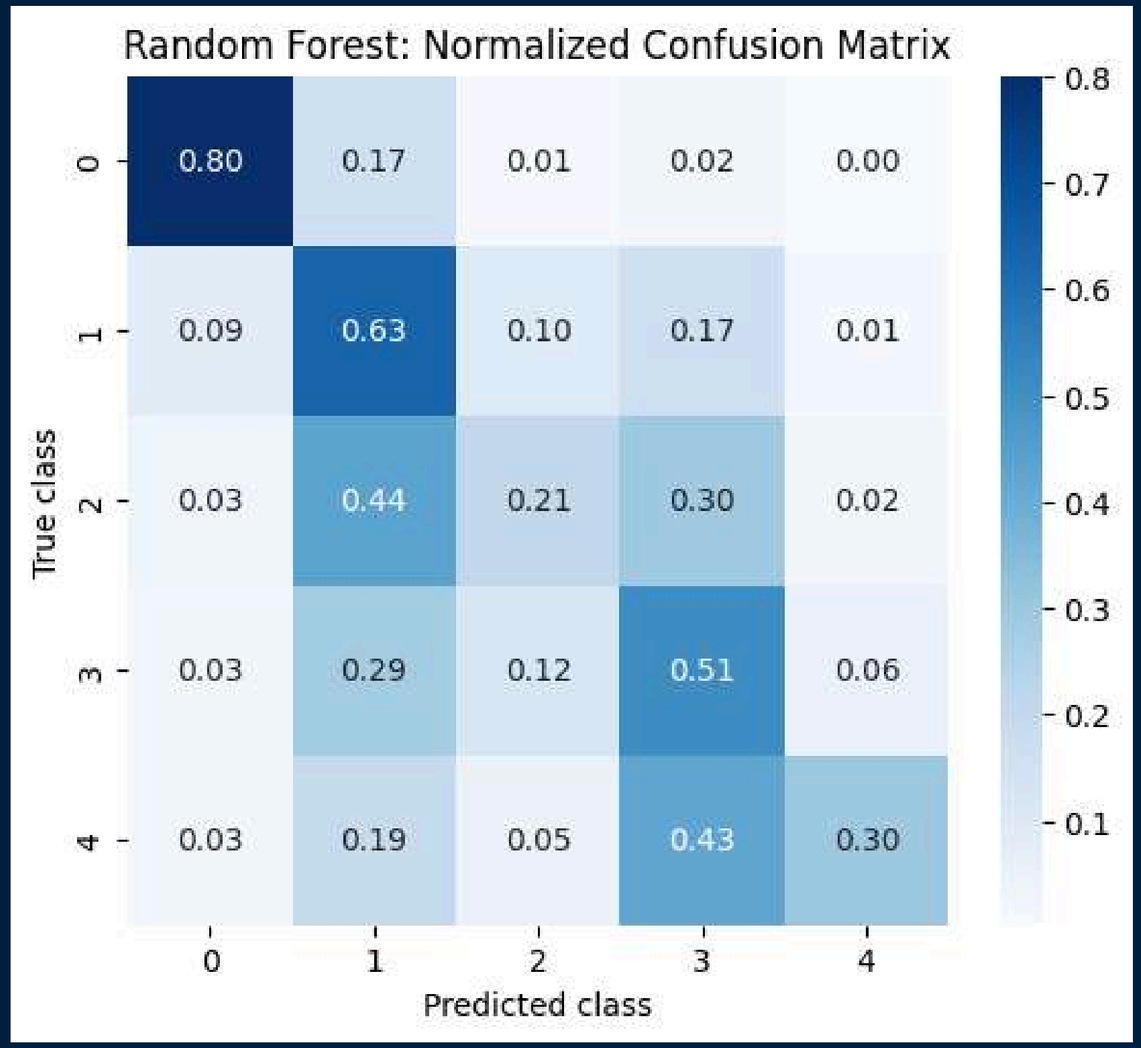
```
Classification report:
              precision    recall  f1-score   support

           0       0.85      0.80      0.82     34481
           1       0.44      0.63      0.52     34489
           2       0.43      0.21      0.28     27613
           3       0.40      0.51      0.45     27507
           4       0.60      0.30      0.40     13783

    accuracy                           0.53    137873
   macro avg       0.54      0.49      0.49    137873
weighted avg       0.55      0.53      0.52    137873

RMSE: 1.0687
QWK (Quadratic Weighted Kappa): 0.6505
```

Confusion Matrix (recall visualization) :

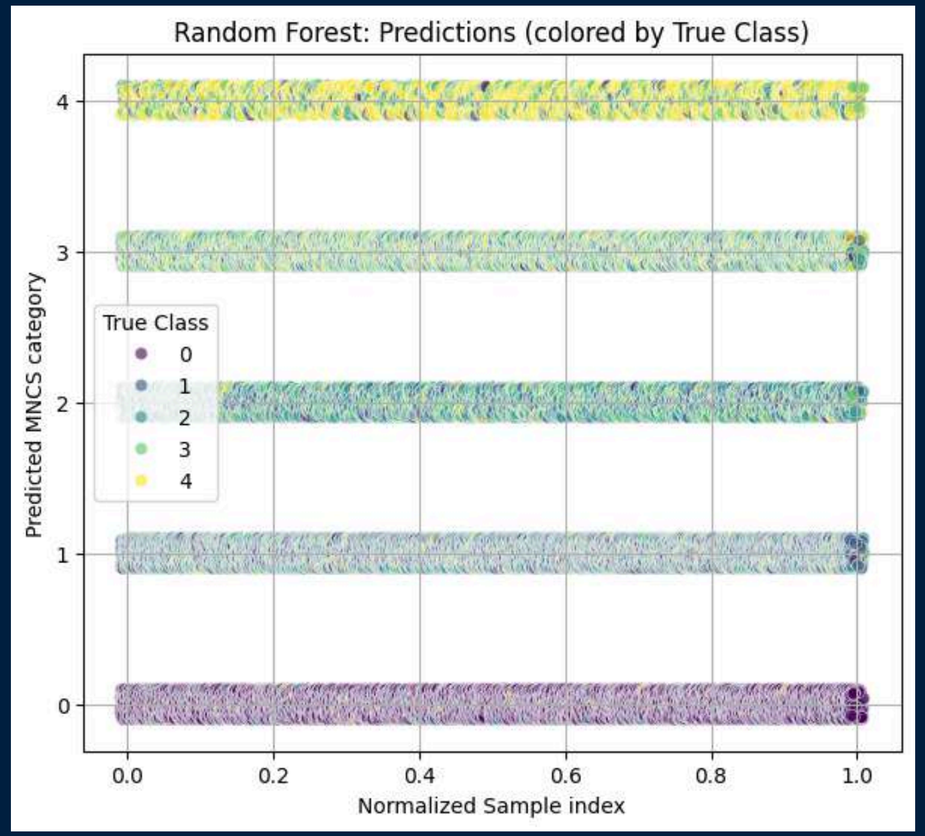


HSDSLab
Human & Social Data Science

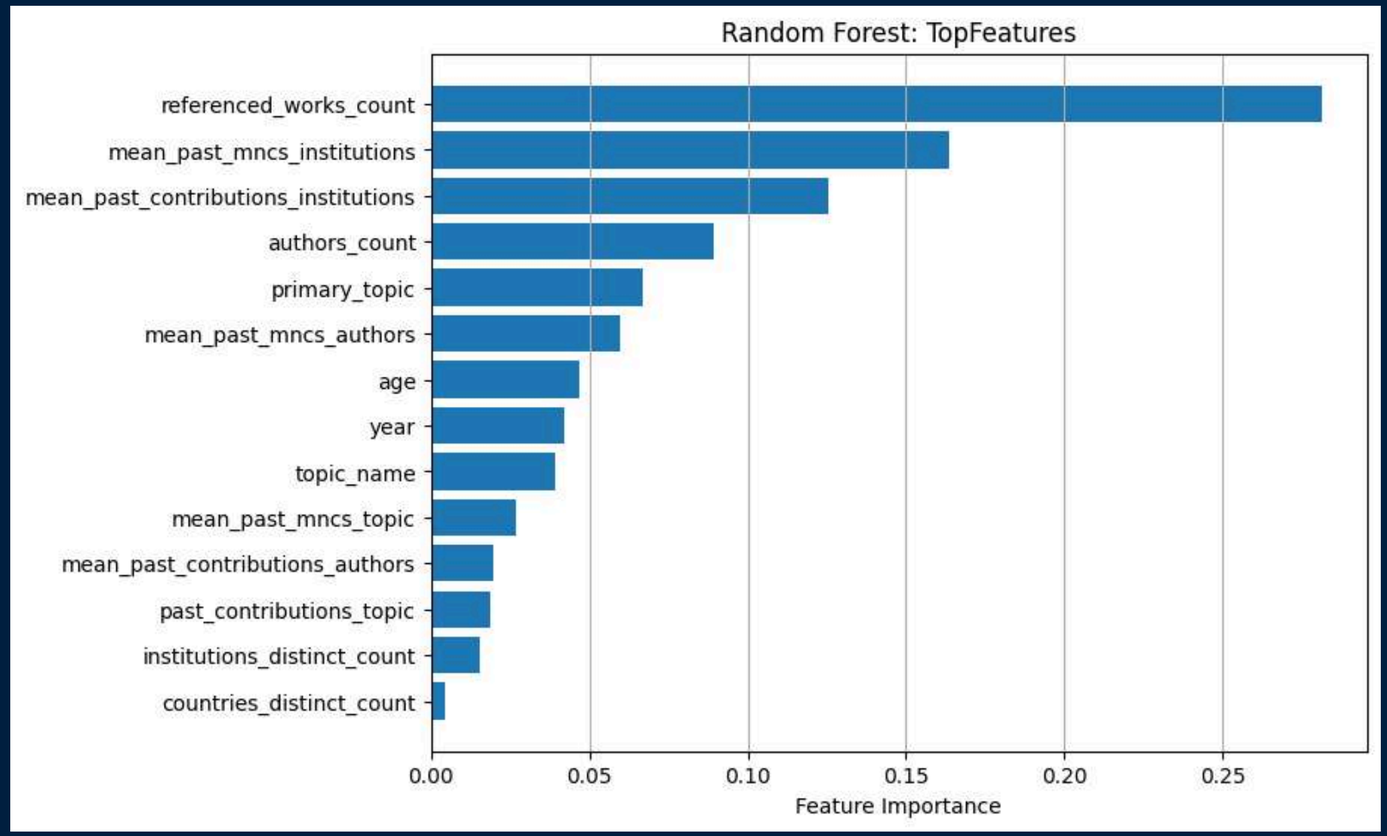
Modelling

General Modelling - Random Forest

A collection of decision trees whose outputs are averaged or voted



Top Features :



HSDSLab
Human & Social Data Science

Modelling

General Modelling - LightGBM

Builds trees leaf-wise in sequence, each correcting errors from previous trees.

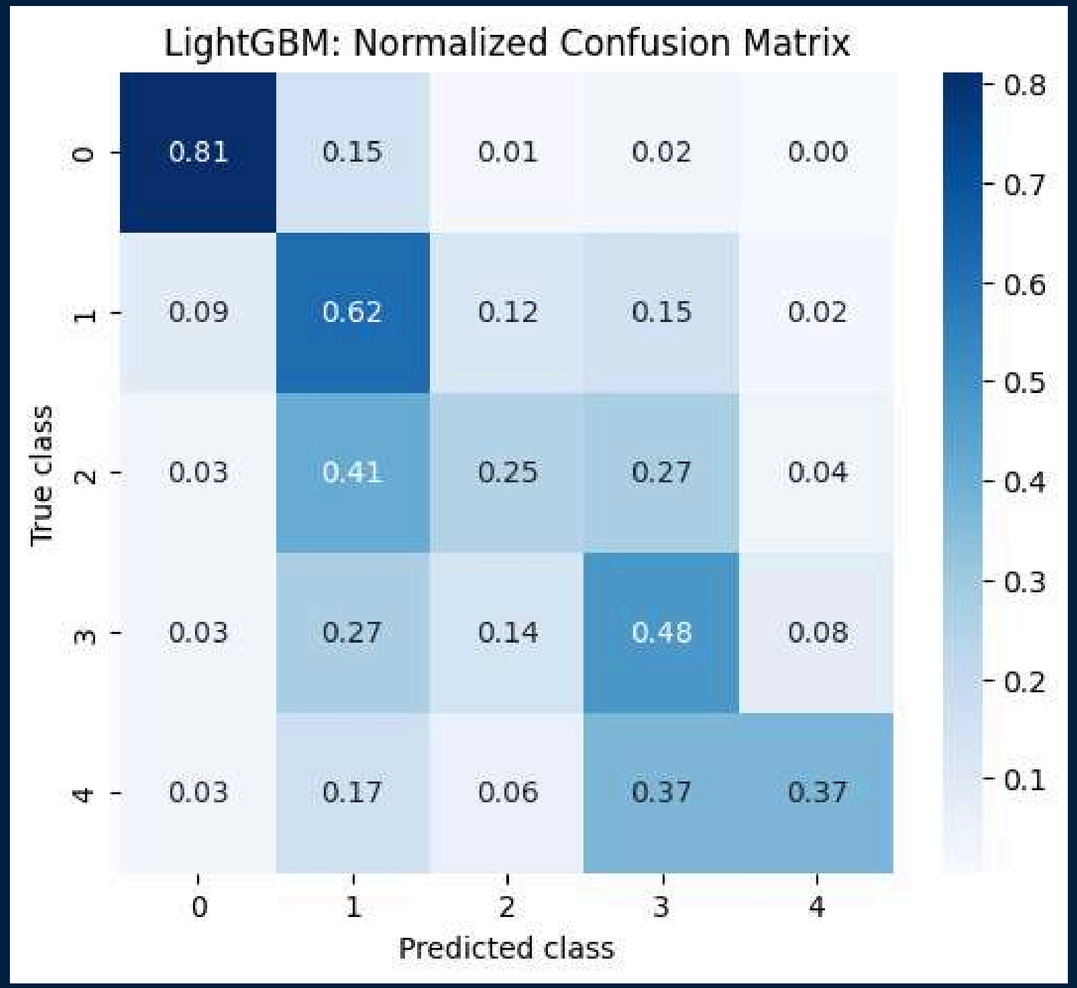
```
Classification report:
              precision    recall  f1-score   support

           0       0.85      0.81      0.83     34481
           1       0.45      0.62      0.52     34489
           2       0.43      0.25      0.31     27613
           3       0.42      0.48      0.45     27507
           4       0.55      0.37      0.44     13783

    accuracy                           0.54    137873
   macro avg       0.54      0.51      0.51    137873
weighted avg       0.55      0.54      0.53    137873

RMSE: 1.0627
QWK (Quadratic Weighted Kappa): 0.6614
```

Confusion Matrix (recall visualization) :

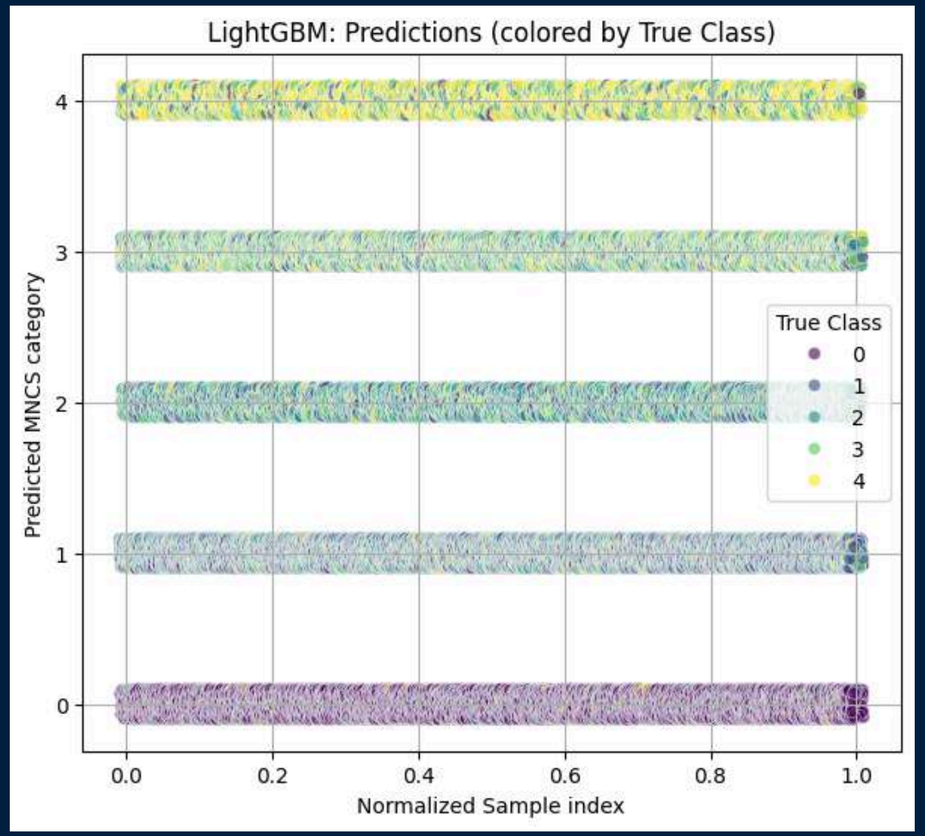


HSDSLab
Human & Social Data Science

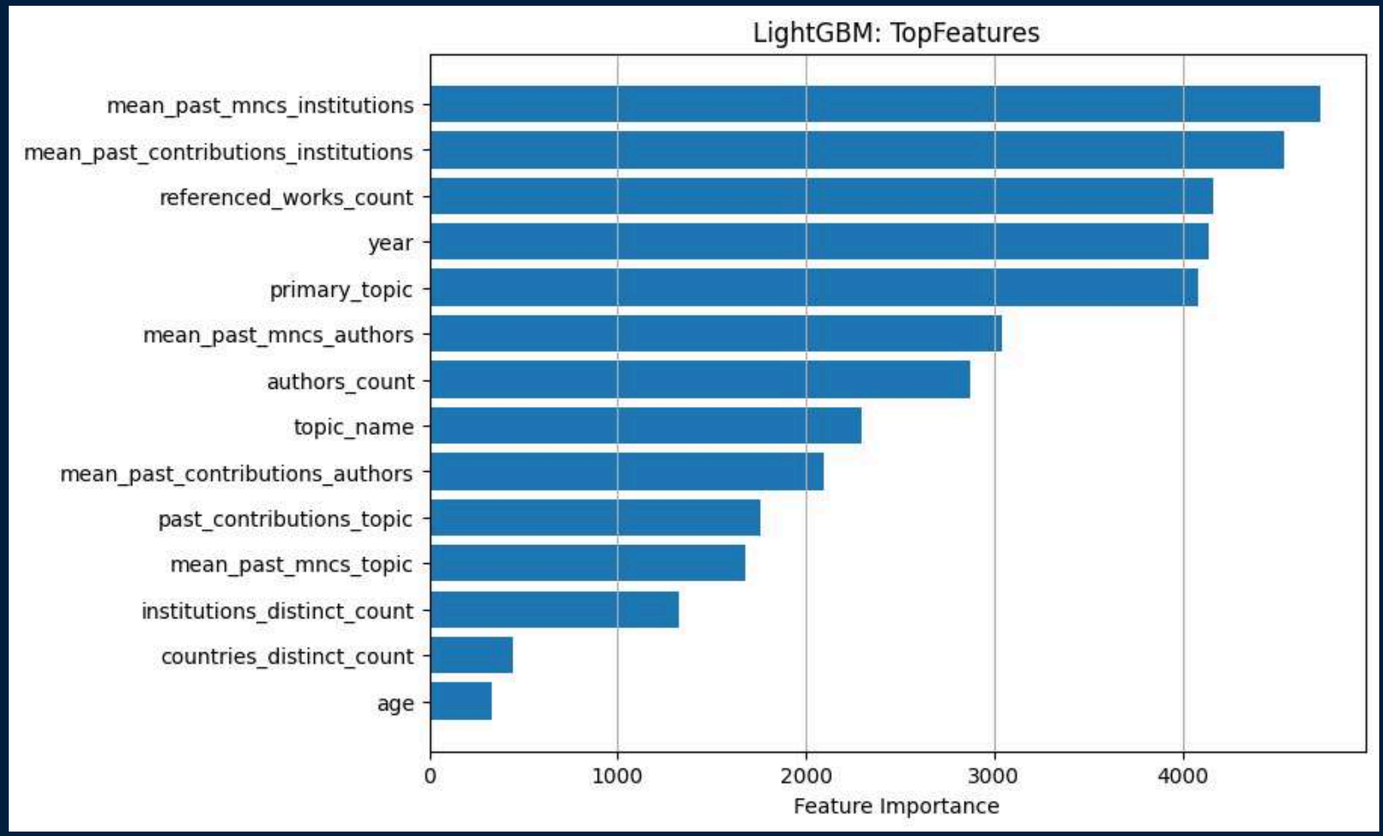
Modelling

General Modelling - LightGBM

Builds trees leaf-wise in sequence, each correcting errors from previous trees.



Top Features :



HSDSLab
Human & Social Data Science

Modelling

General Modelling - MLP

A network of interconnected neurons that transforms inputs through hidden layers to predict outputs

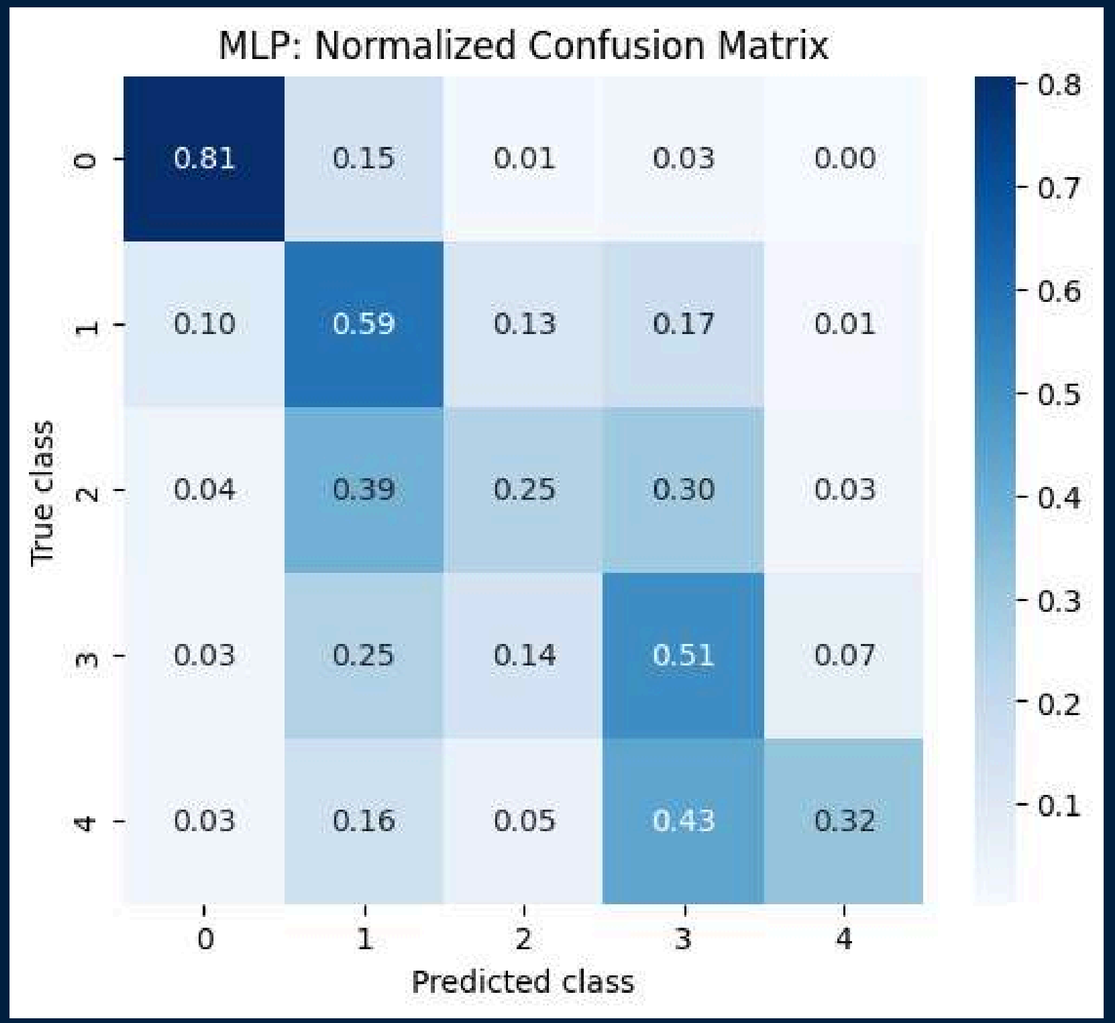
```
Classification report:
              precision    recall  f1-score   support

           0       0.83      0.81      0.82     34481
           1       0.44      0.59      0.51     34489
           2       0.42      0.25      0.31     27613
           3       0.40      0.51      0.45     27507
           4       0.58      0.32      0.41     13783

    accuracy                           0.53    137873
   macro avg       0.53      0.49      0.50    137873
weighted avg       0.54      0.53      0.52    137873

RMSE: 1.0650
QWK (Quadratic Weighted Kappa): 0.6574
```

Confusion Matrix (recall visualization) :

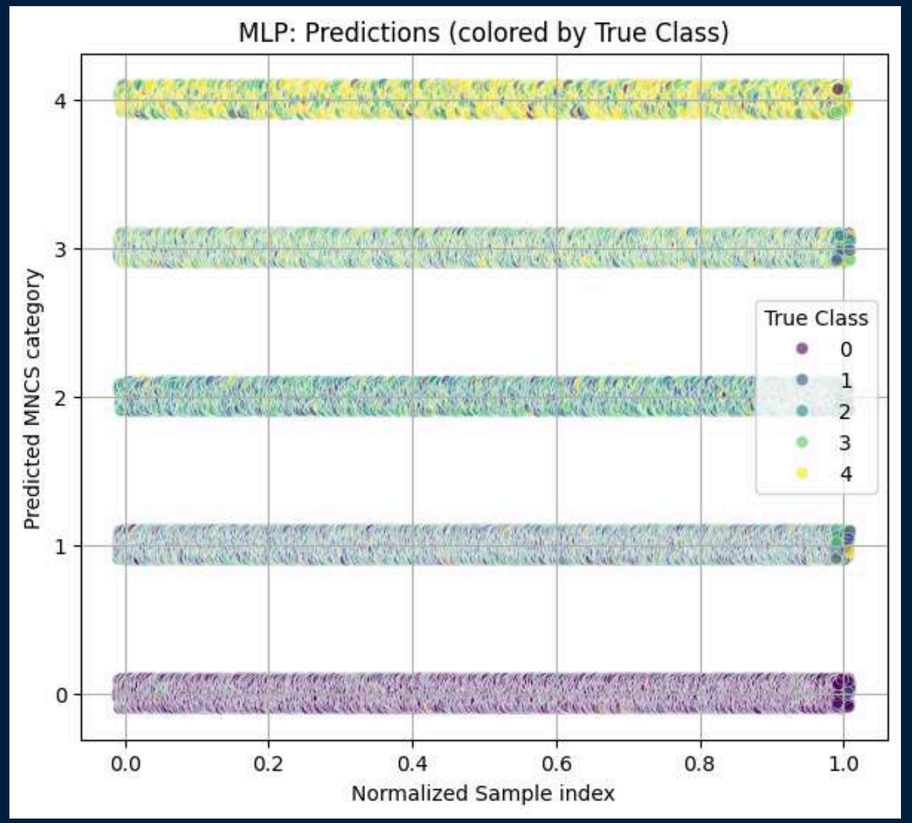


HSDSLab
Human & Social Data Science

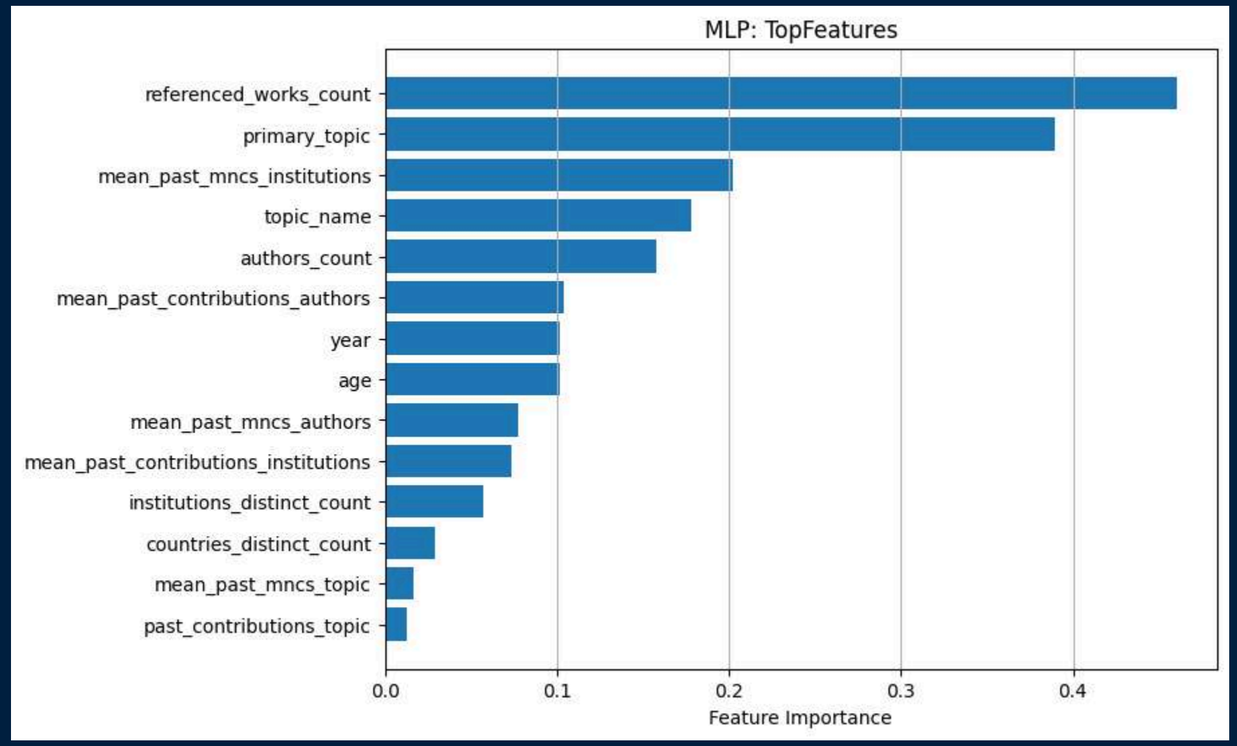
Modelling

General Modelling - MLP

A network of interconnected neurons that transforms inputs through hidden layers to predict outputs



Top Features :



HSDSLab
Human & Social Data Science

Modelling

General Modelling - XGBoost

Builds trees sequentially, each learning from the mistakes of the previous ones like LightBGM

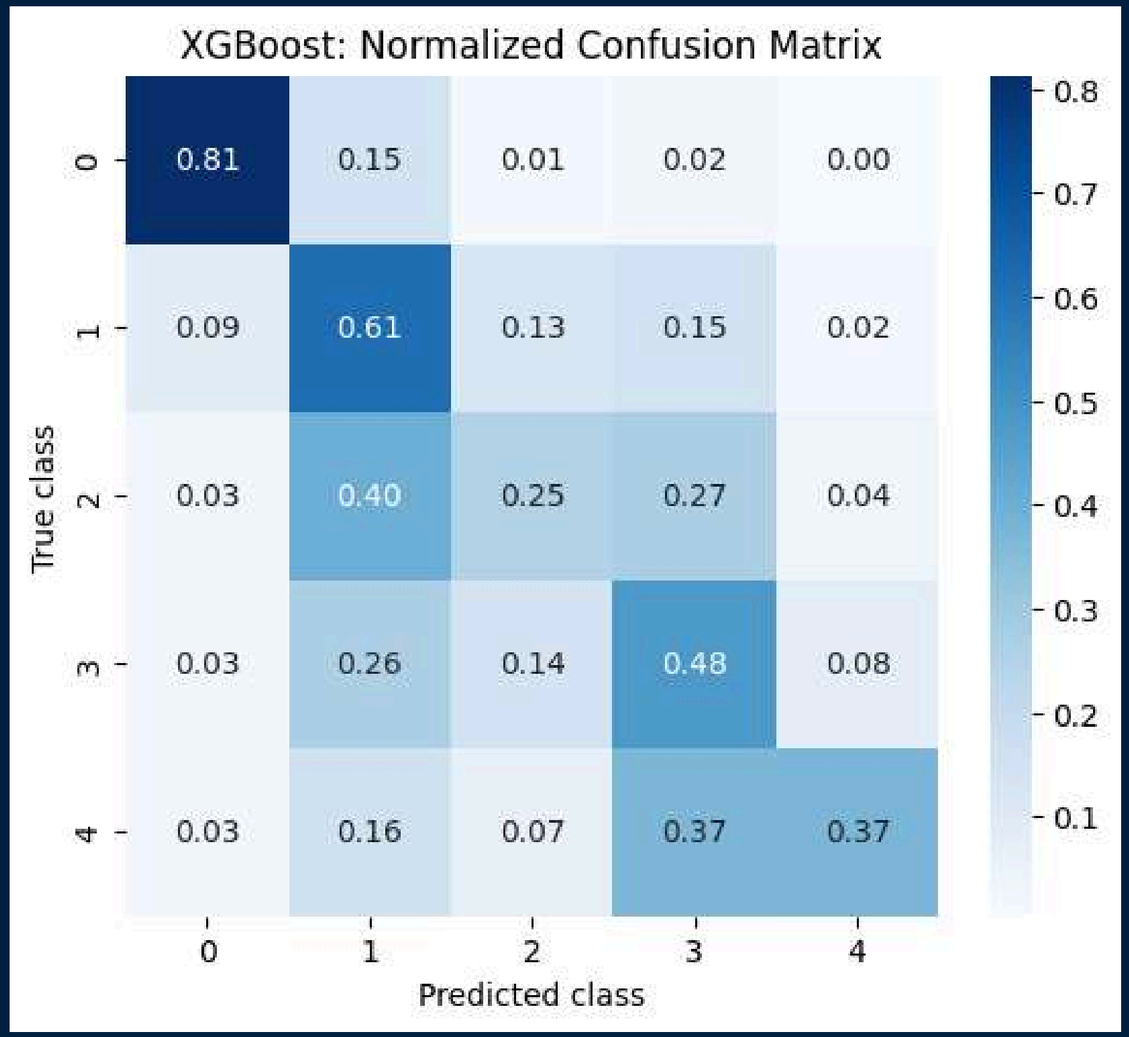
```
Classification report:
              precision    recall  f1-score   support

           0       0.85      0.81      0.83     34481
           1       0.45      0.61      0.52     34489
           2       0.42      0.25      0.32     27613
           3       0.42      0.48      0.45     27507
           4       0.55      0.37      0.45     13783

    accuracy                           0.54    137873
   macro avg       0.54      0.51      0.51    137873
weighted avg       0.55      0.54      0.53    137873

RMSE: 1.0555
QWK (Quadratic Weighted Kappa): 0.6657
```

Confusion Matrix (recall visualization) :

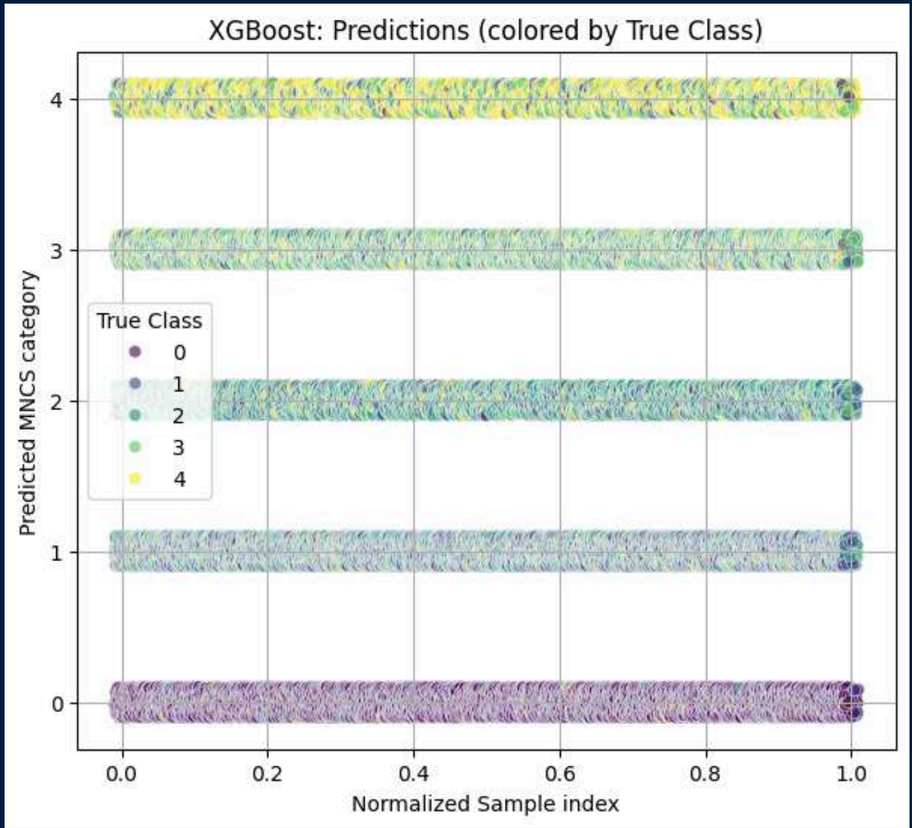


HSDSLab
Human & Social Data Science

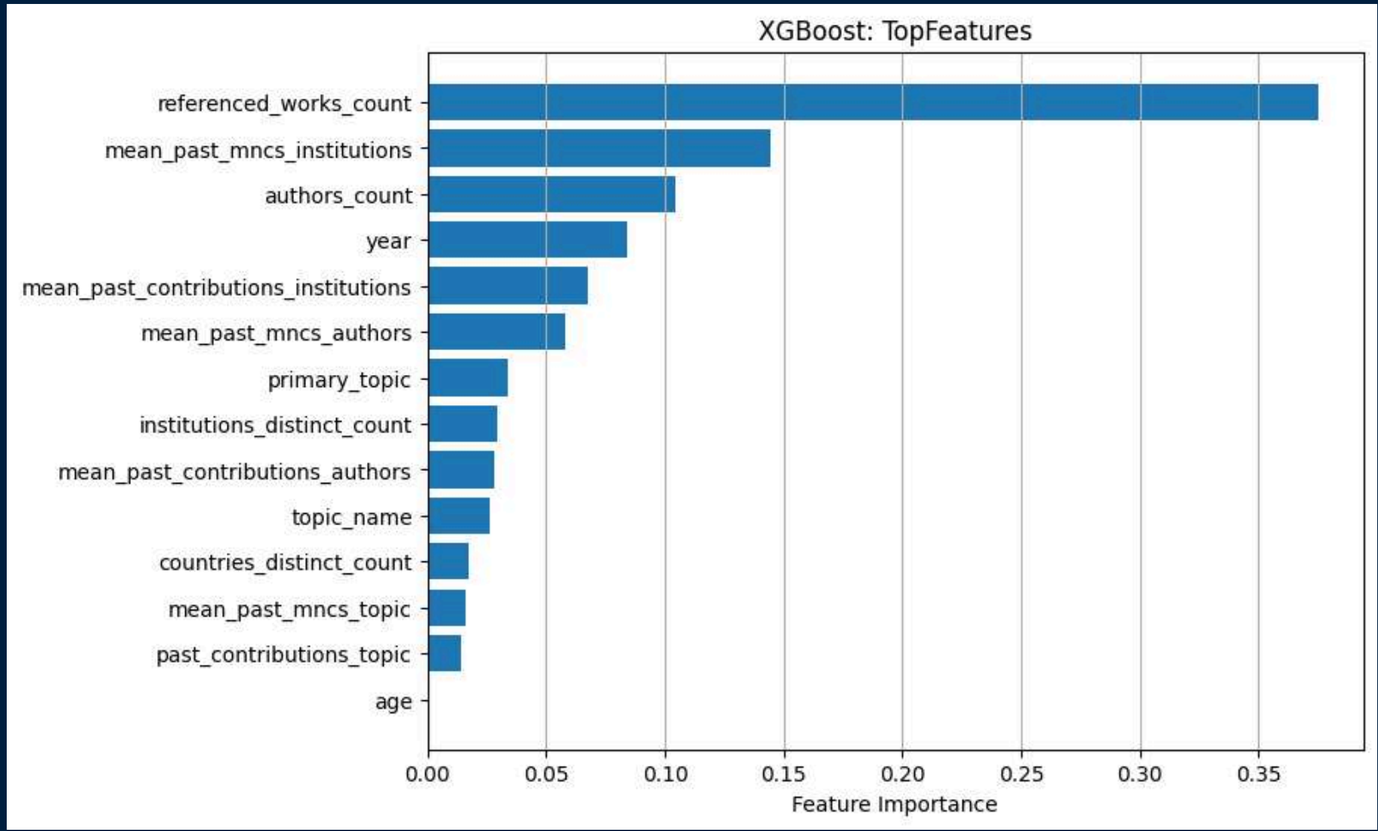
Modelling

General Modelling - XGBoost

Builds trees sequentially, each learning from the mistakes of the previous ones like LightGBM



Top Features :



HSDSLab
Human & Social Data Science

Modelling

General Modelling - Conclusion

The higher the accuracy/qwk, the better
The lower the rmse, the better

Final model performance:		qwk	rmse
	accuracy		
Logistic Regression	0.464166	0.547114	1.220659
Random Forest	0.532577	0.65048	1.068749
XGBoost	0.541223	0.665747	1.055515
LightGBM	0.540642	0.661411	1.06273
LDA	0.418022	0.469761	1.272586
kNN	0.444561	0.535006	1.22556
GaussianNB	0.364843	0.429004	1.262751
MLP	0.531373		

- XGBoost is the best-performing model
- For most models, the top feature is the number of referenced works
- Mean past insitution metrics can also be important (especially for LightGBM)
- Primary Topic is not negligible (LDA, KNN, MLP)
- Models tend to predict the bottom category more accurately

Conclusion

Conclusion

BJP Among Q1 Journals

- BJP is among the top journals in pharmacology, including Q1 journals.
- However, recent decline in publication rate may reduce its visibility and citations
- Citations are more consistent across publications

- Most publications come from British and European authors and institutions
- Popular topics are similar to other top journals, but BJP focuses more on some areas

BJP : From the UK, Leading Pharmacology Worldwide

HSDSLab
Human & Social Data Science

Conclusion

Global Analysis of Journal Publications

- Increasing number of publications : The journals are publishing more articles over time.
- Growing popularity and reach : Publications are being cited and shared more widely.
- More diverse authorship : Increasing variety in authors, countries, and institutions represented.
- Greater diversity in topics : Research covers a broader range of subjects.
- Number of authors, countries, and institutions has a slight impact on popularity : Statistical tests show only a modest effect on citations.
- Collaboration patterns matter : Some Asian countries publish a lot but collaborate less, resulting in fewer citations.
- Topic popularity varies over time : Certain topics are more cited than others, with trends changing across decades.
- Modelling confirms links : Citation prediction models perform better than random, showing relationships.

HSDSLab
Human & Social Data Science

Conclusion

General Works Analysis

Some Precisions

- Most plots shown – Additional visualizations exist but are not included in this presentation
- Interactive plots created – Some visualizations allow dynamic exploration of the data (topic modeling)
- Videos illustrate trends – Evolution of top topics, authors, and other metrics over time.

Possible Improvements ?

- Enhance existing plots – Improve clarity and add more details to current visualizations.
- Deeper (topic) modeling – Explore topics, models more in greater depth.
- Citation prediction tool : Develop a tool to predict the number of citations for new publications based on key metrics.

HSDSLab
Human & Social Data Science

Thank you!

HSDSLab
Human & Social Data Science