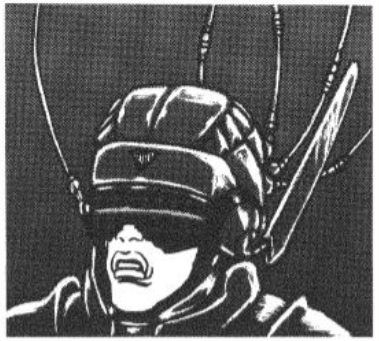
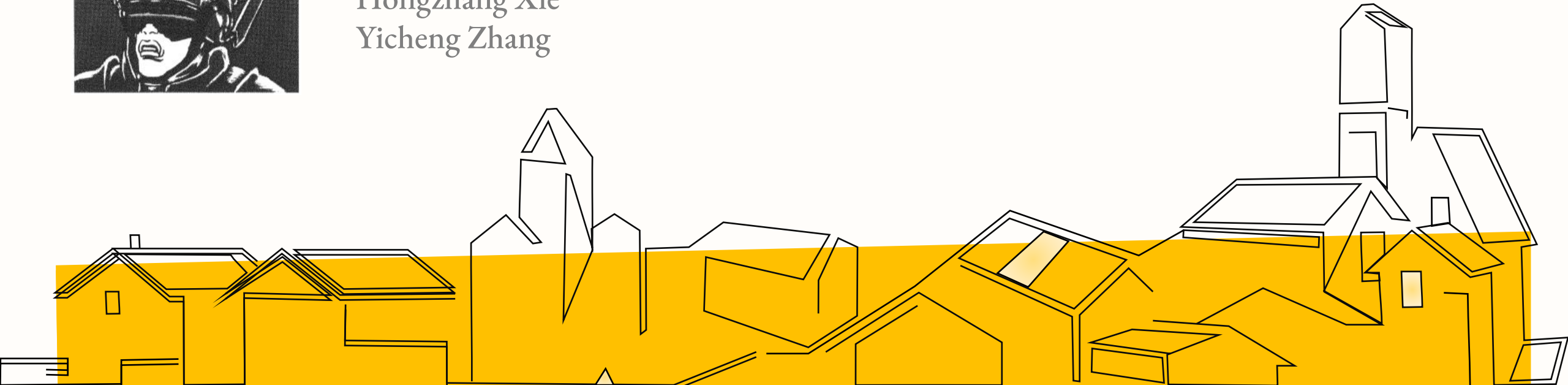
*MACS30122 Final Group Project*
*University of Chicago*

# Portrait of Economic Academia: Evidence from Top 20 Economics Journals

Collaboration Network & Topic Modeling Analysis

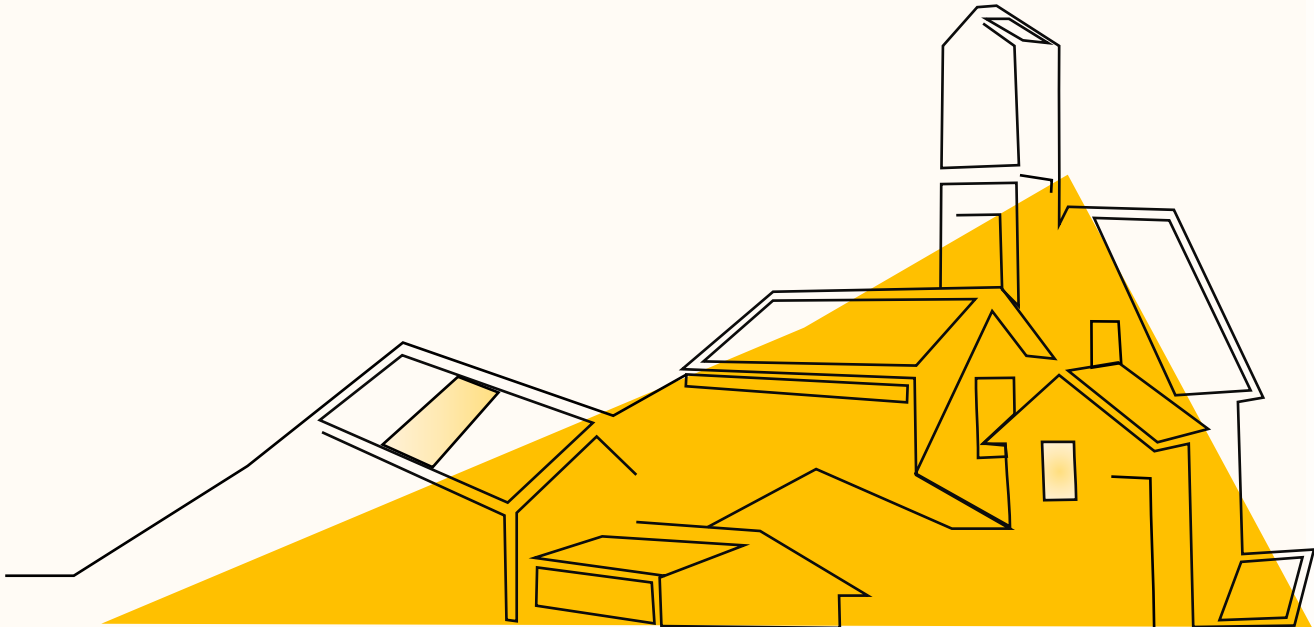



*Netrunners*
Jerry Cheng
Hongzhang Xie
Yicheng Zhang

# Table of Contents



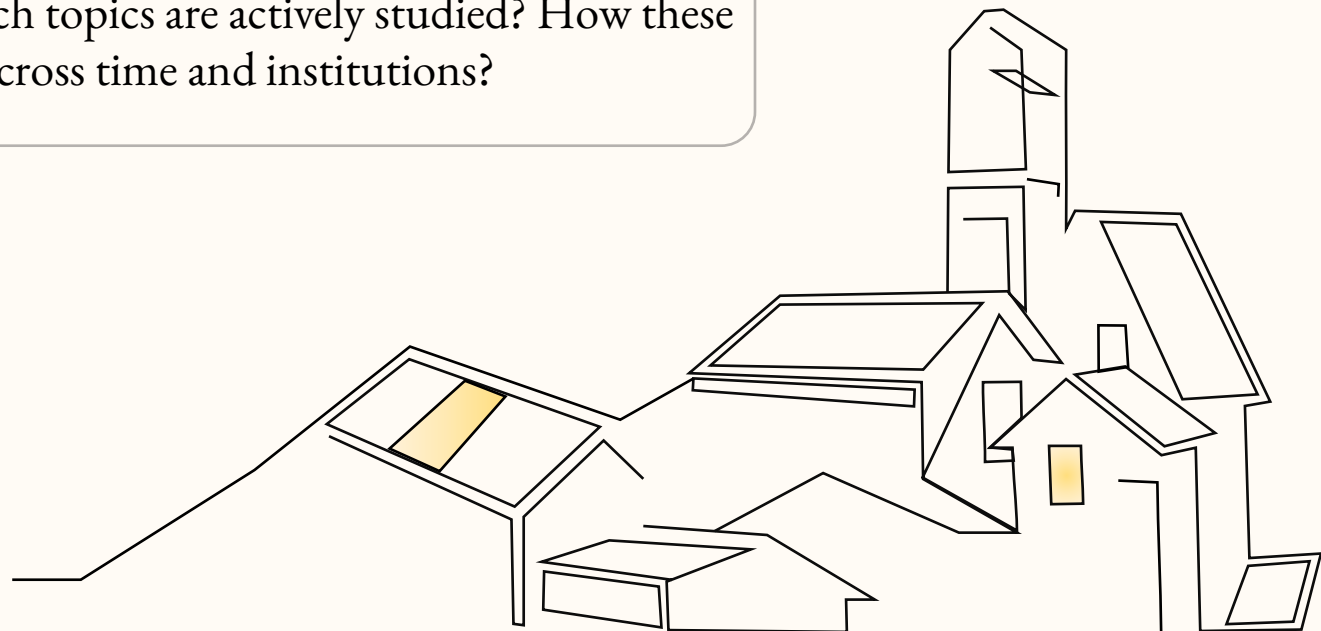


# Research topic:

## Examine the characters of the economic research collaboration structure

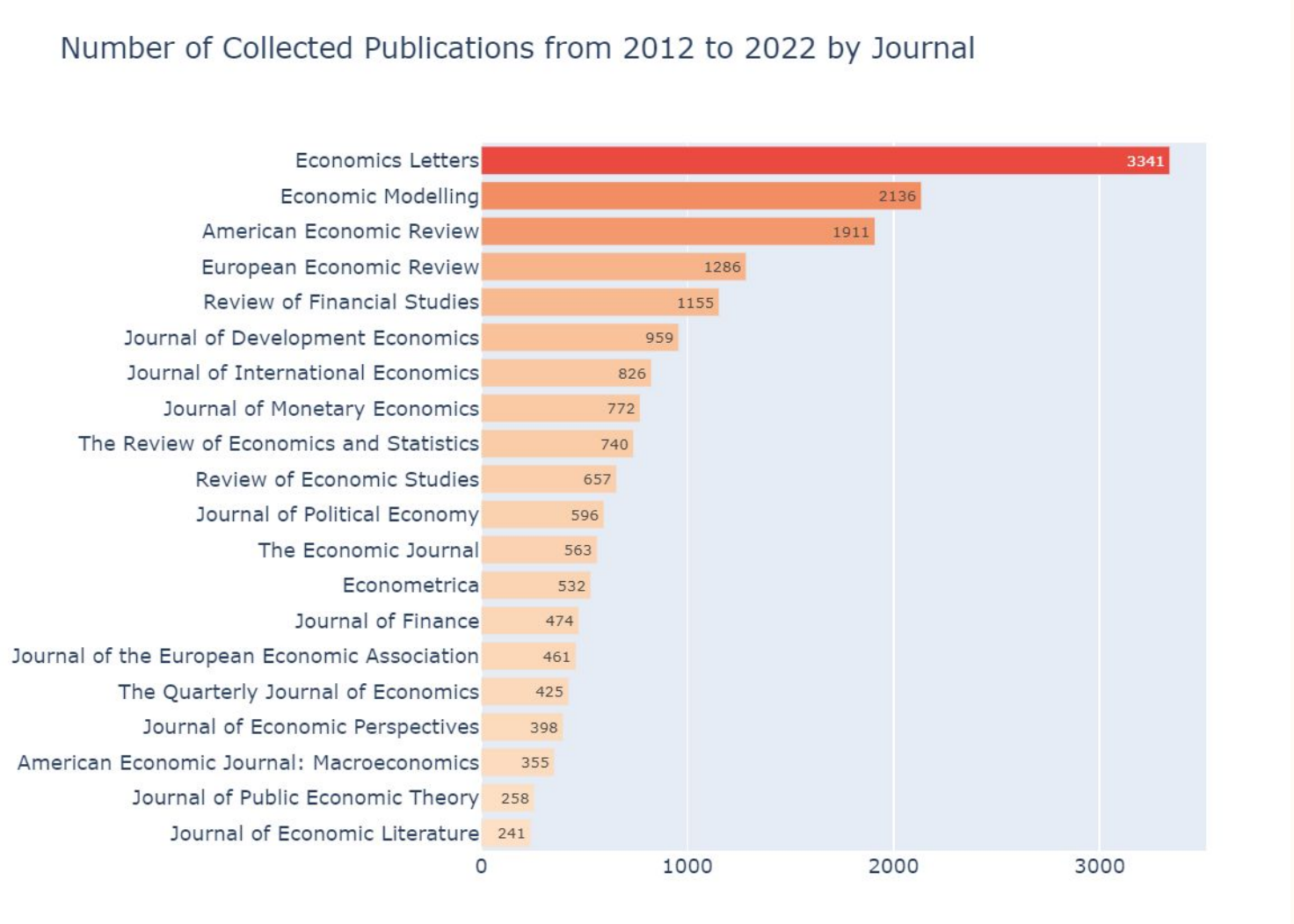
| | |
|---|---|
| **Question 1** | What are the network features and collaboration patterns at the institutional level? |
| **Question 2** | What economics research topics are actively studied? How these topics vary and evolve across time and institutions? |

Data Description



# Data Description -Top 20 Economics Journals Data

- We collected the data of top 20 economics journals ranked by h5-index (2017 – 2021).
- Our raw data have 79,565 publications from 1886 till now.
- We select 18,086 publications from 2012 to 2022 as the research sample.
- We have three databases:
  - ➢ Top 20 Economics Journals Data
  - ➢ Google Scholar Author Data
  - ➢ Authors' Affiliations Data



# Data Description
## Top 20 Economics Journals Data

## Data sources

### Official Websites:

- American Economic Association
- Oxford Academic
- Science Direct

### Third-party database :

- ideas.repec.org
- econpapers.repec.org

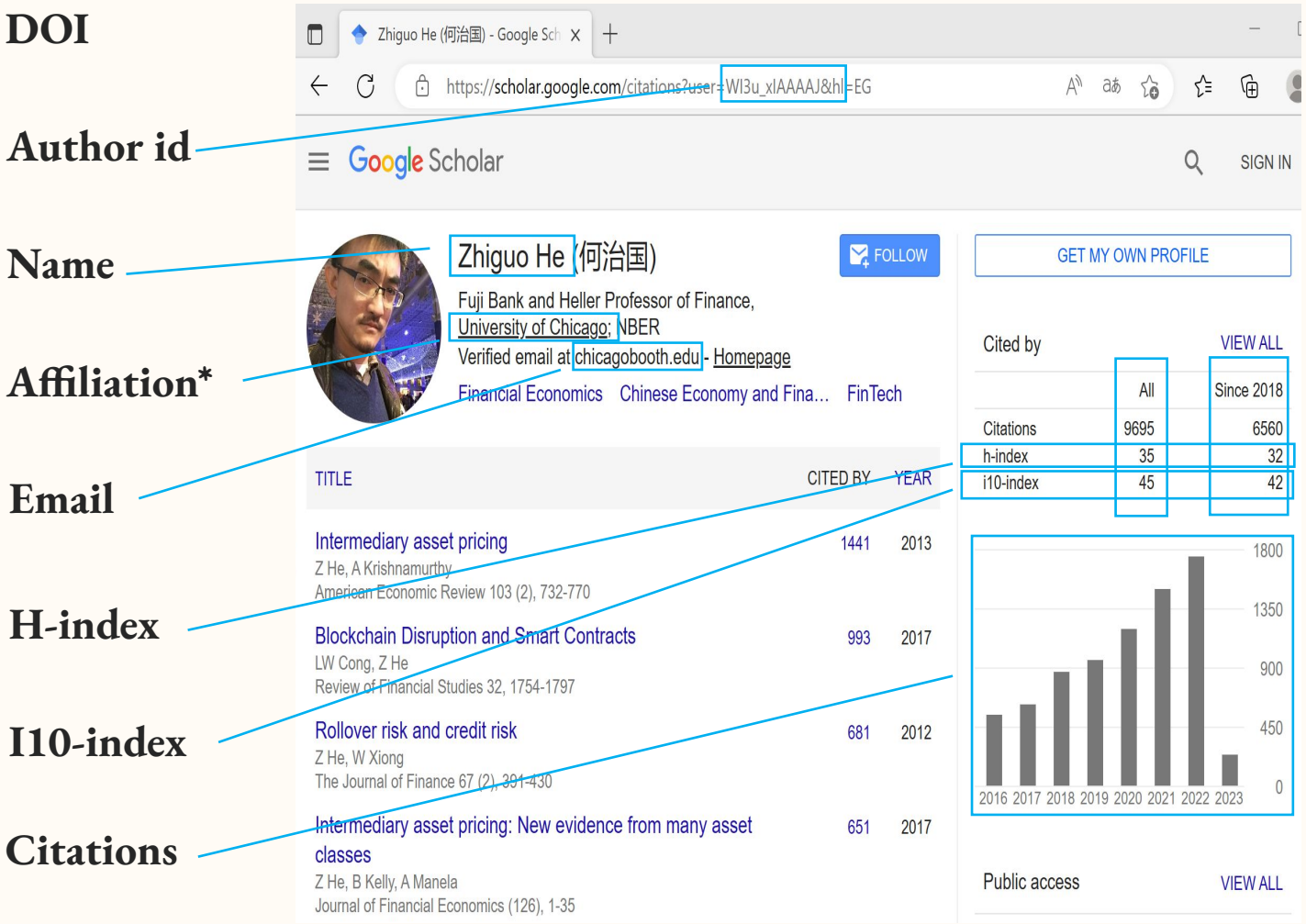
# Data Description
## Google Scholar Author Data

### Links of two databases

- We input the **DOI in Top 20 Economics Journals Database** to search the **Authors** of the publications in Google Scholar.

### Data sources

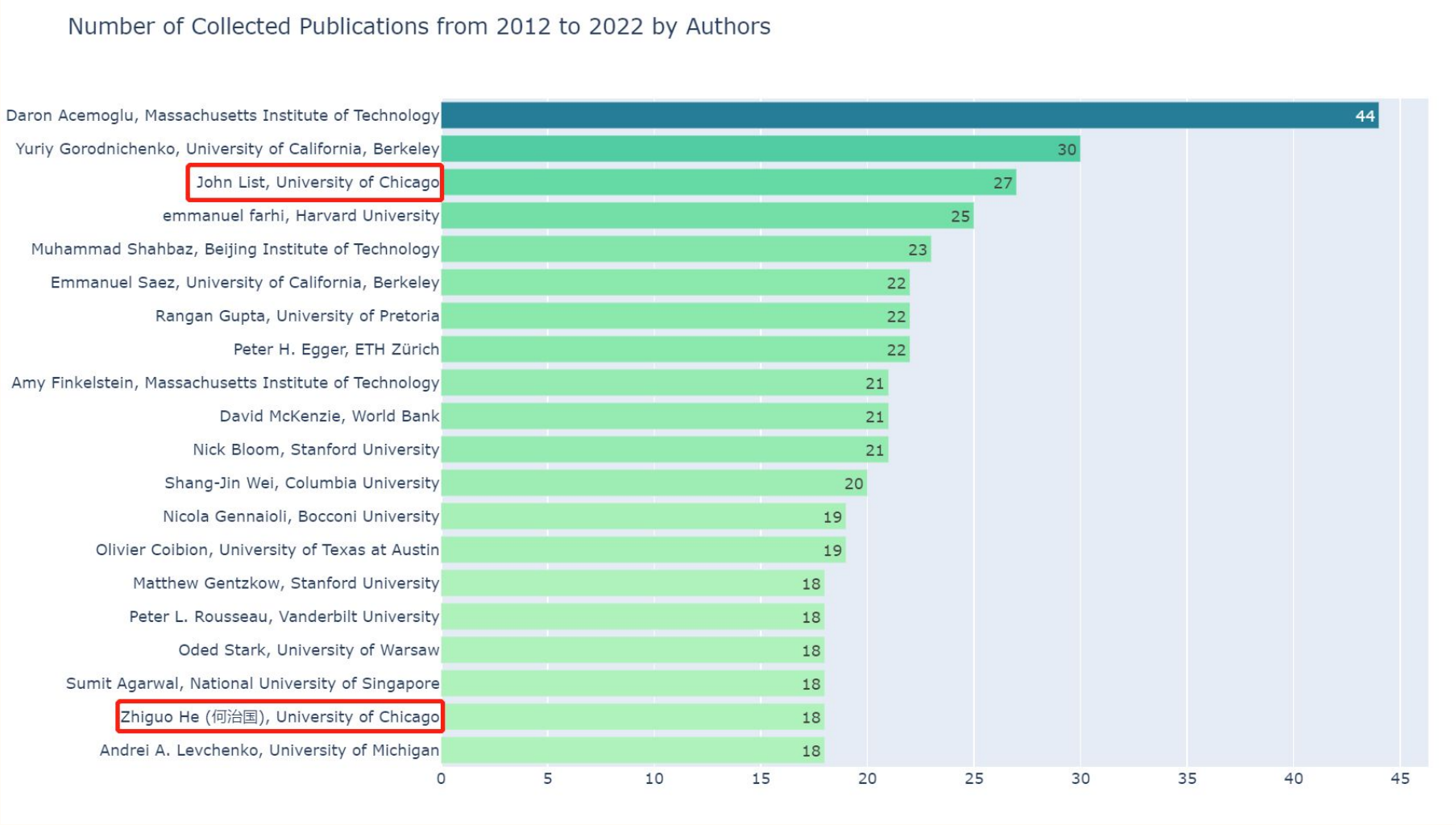
- Google Scholar Author Database
- Semantics Scholar Database (For incompleteness data)




* If the author has **Multiple Affiliations**, we use the **Affiliation of the Email.**

# Data Description - Google Scholar Author Data

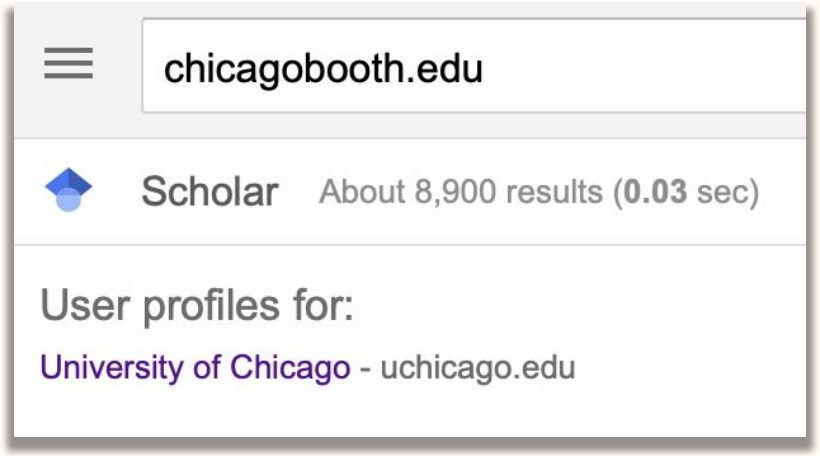
- Our raw data have 19,585 records of authors
- We select 13,654 authors who have publications from 2012 to 2022 as the research sample.



# Data Sources and Description - Authors' Affiliations Data

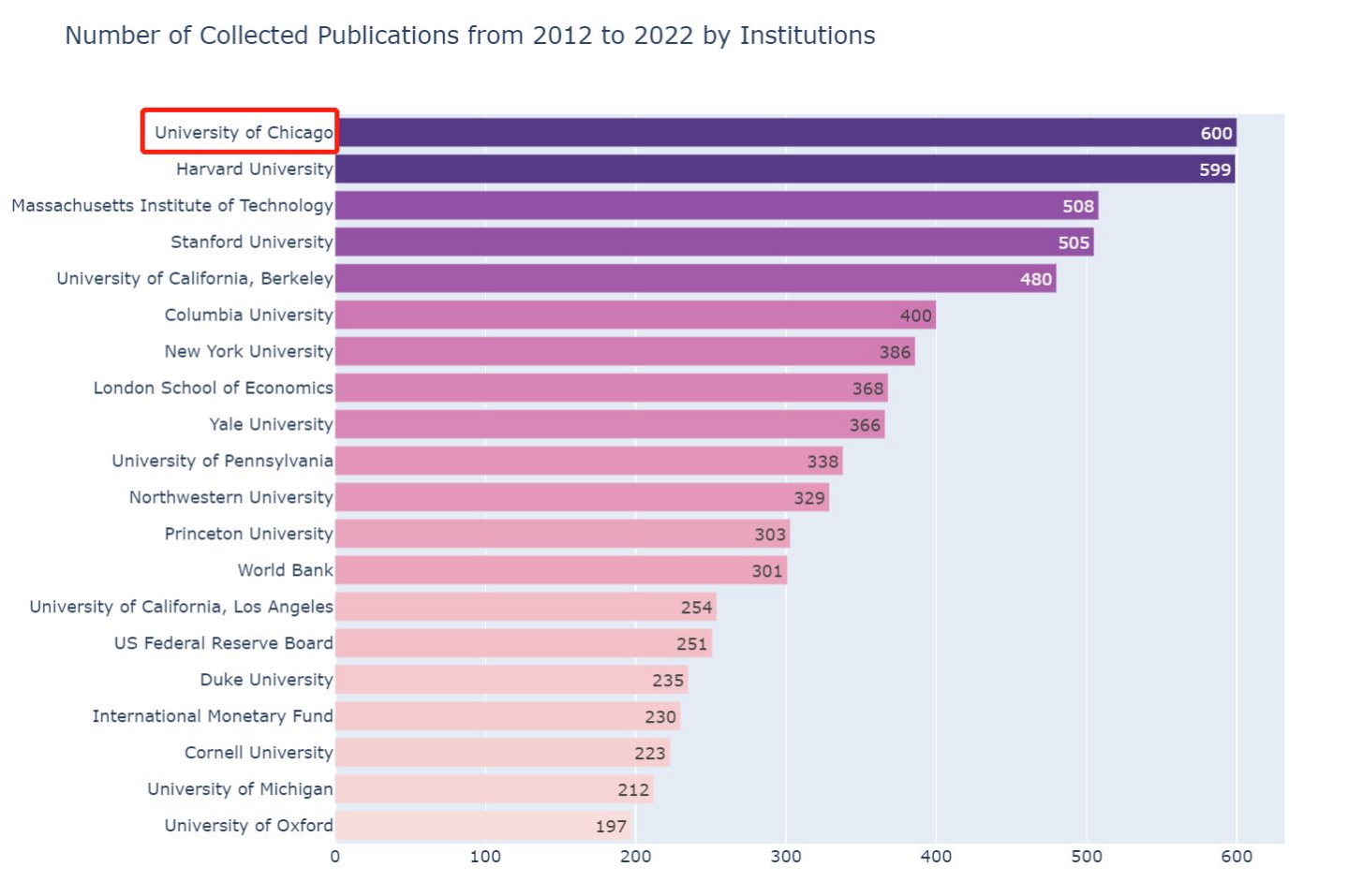
## Data sources

- Google Scholar Institutions Information

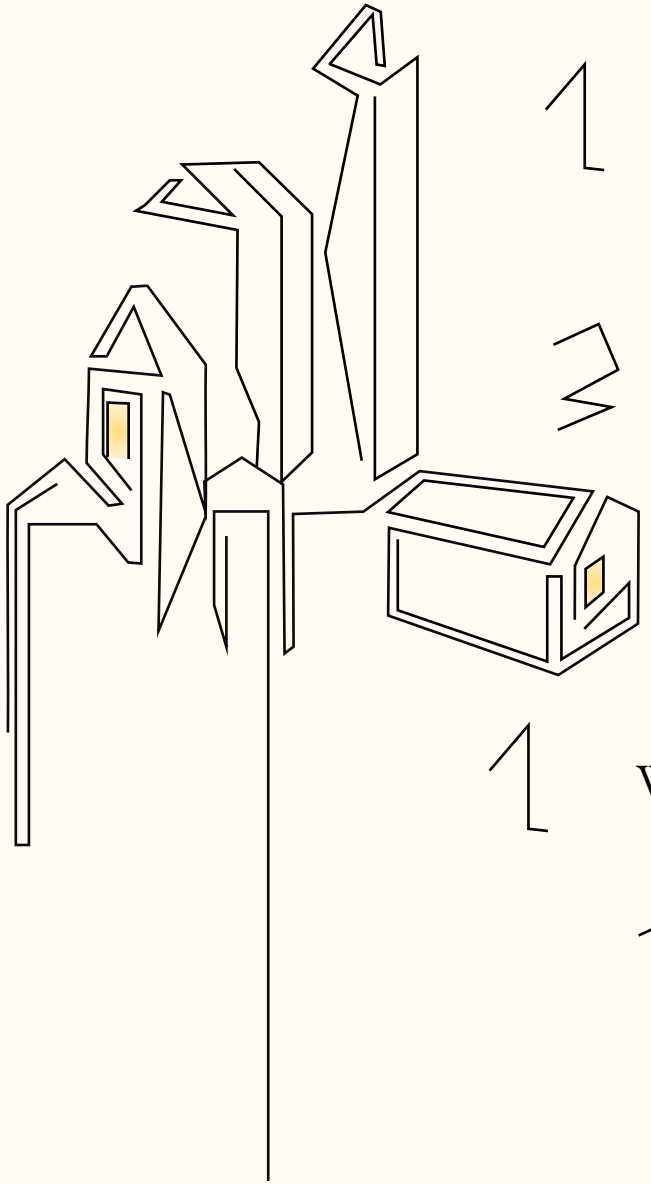



## Data description:

- Our raw data have 1,392 records of institutions.
- We select 1,329 institutions that have authors who publish Top articles from 2012 to 2022 as the research sample.



# Advantages

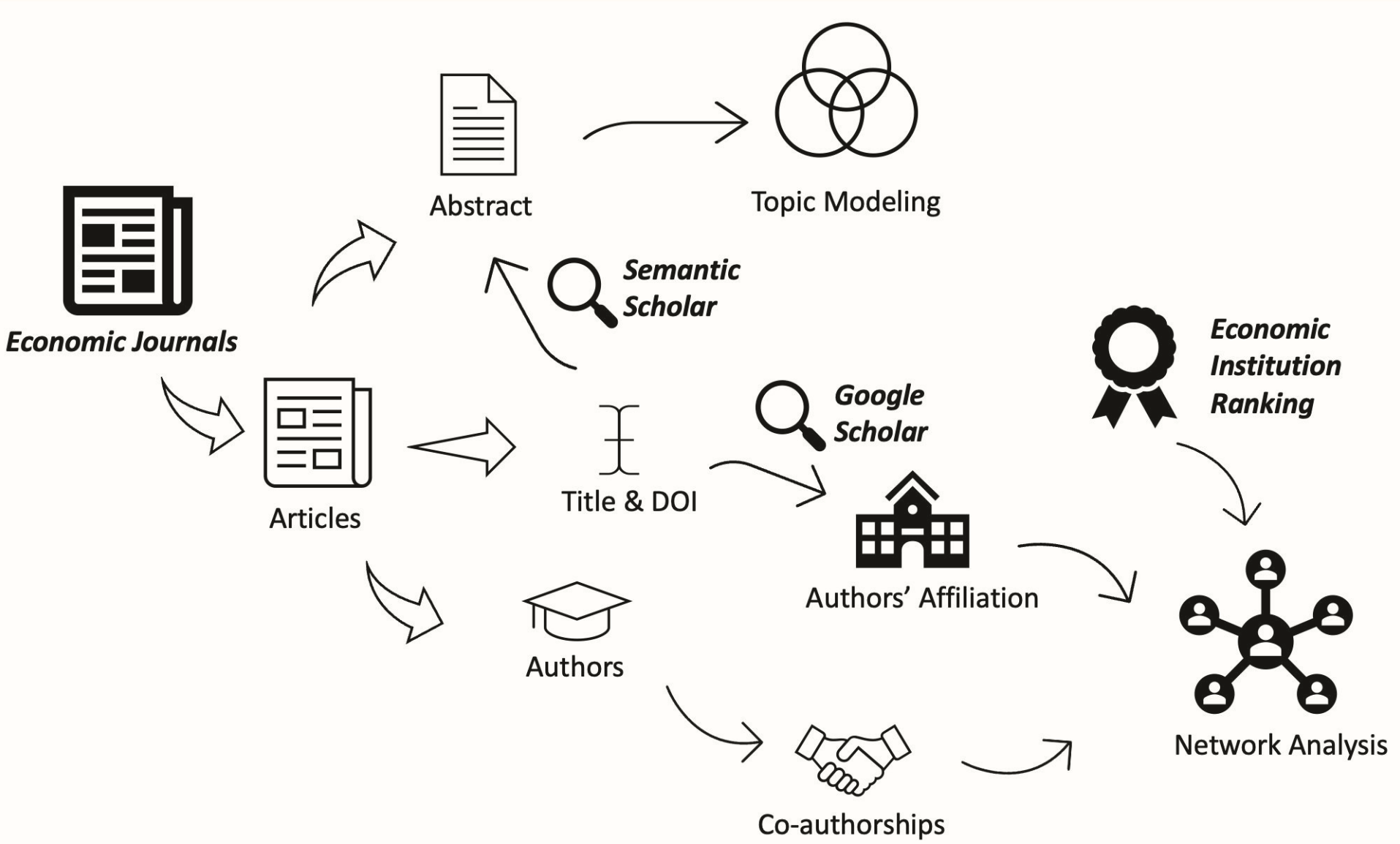


1. Large sample size with multiple attributes.
2. High data quality,
   Crawl data from authoritative databases
3. Innovation, few studies have used similar data sources.
4. Importance, analyzing those data can help us better understanding the research collaboration mechanisms, and improve efficiency.

# Limitations

1. We do not analyze the publication information outside the top 20 economic journals.
2. There are still some missing data in Publish time, DOI, and Authors in the publications before 2000.
3. We did not crawl the data for the authors who do not have Google Scholar profiles.

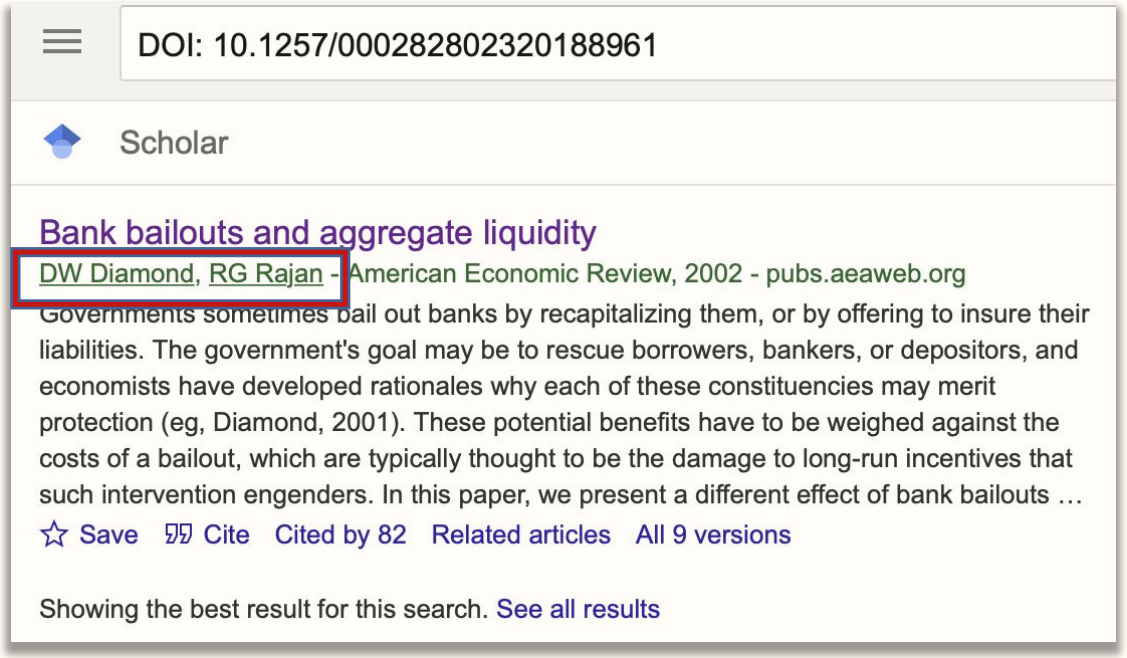
# Data Collection Roadmap



# Building Authors Database

## (1) Retrieve Authors' Google Scholar Page

- Iterate through the article database. For every article, search the article's DOI (if available) in Google Scholar.
- Since the DOI is unique for every article, the researching result is either one or none.
- Scholars that has their own Google Scholar page will have the links embedded in their names.
- Collected *all the embedded links* to revisit later.



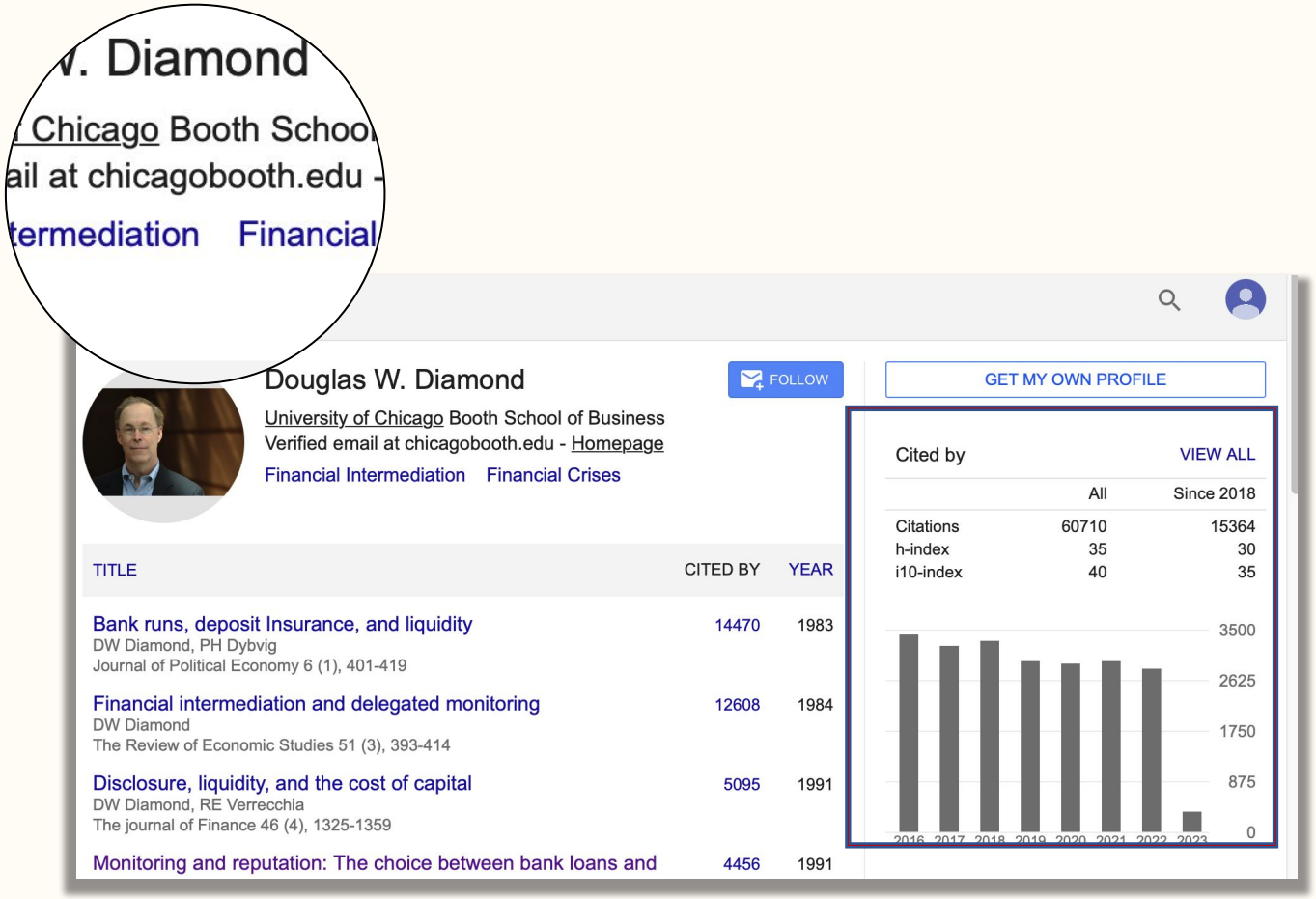
# Building Authors Database

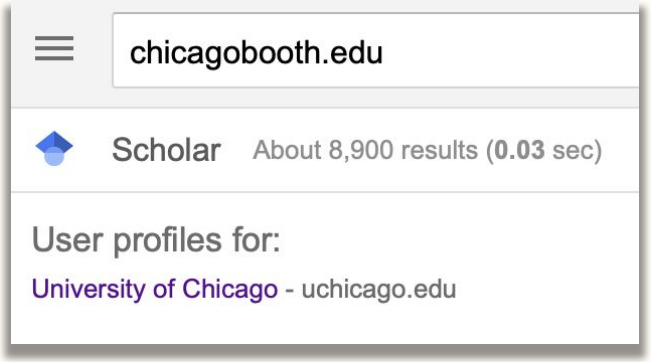
(2) Collect author's info from Google Scholar page

- Revisit the embedded URLs collected.
- Collected information includes
  - the *email address* for this scholar.

  (Google Scholar affiliation section is messy)
  - Citations by year
  - h-index and i10-index listed

- Finally, search the email address collected directly in the Google Scholar.
- The affiliation match with this address will appear in the top box.



# Building Institution Database

Top 10% Economic Institutions (Last 10 Years Publications), as of January 2023 (with details)

Explanations | Ranking | More

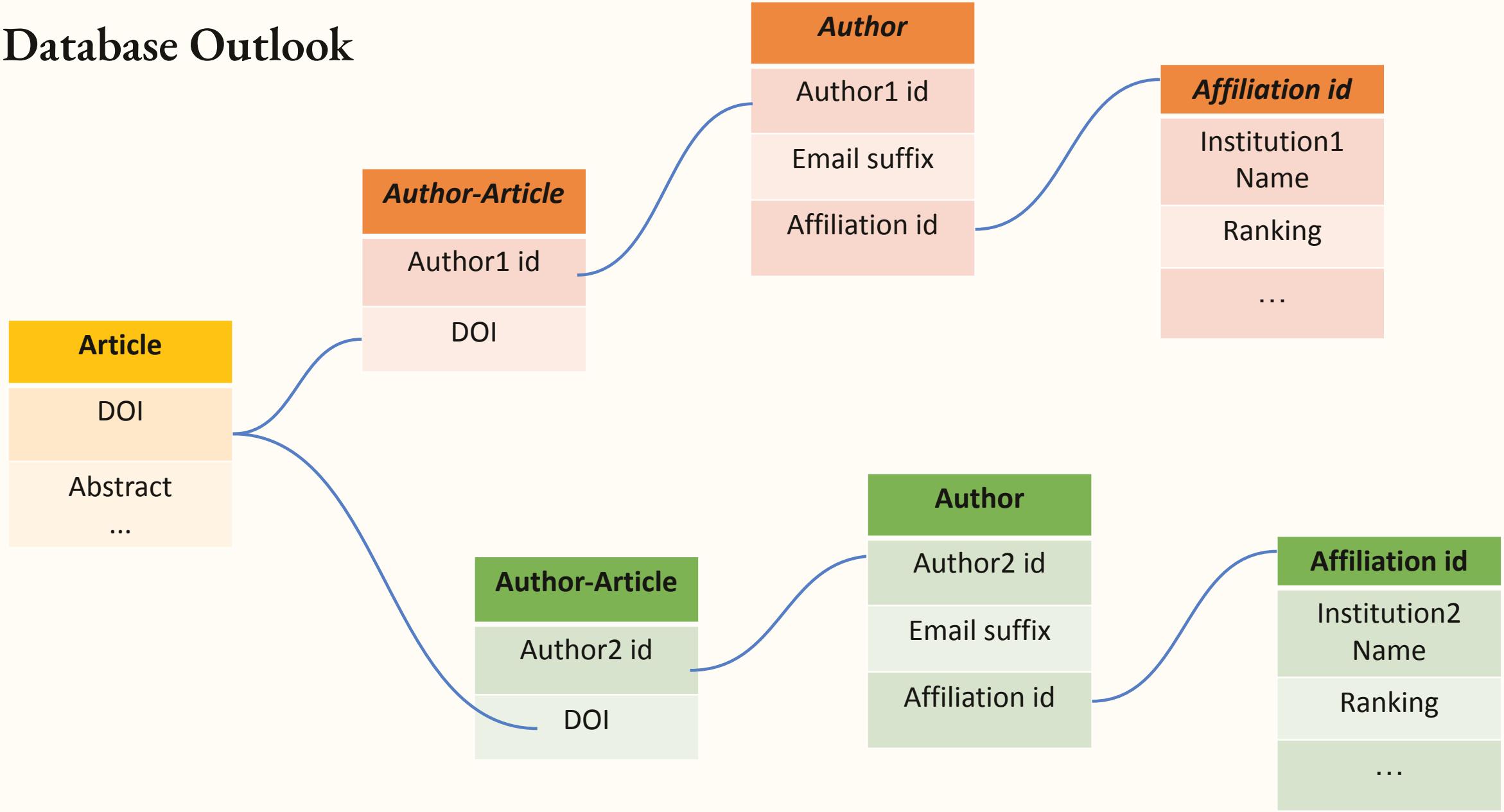
The rankings

| Rank | Institution | Score | Nb Works | DNb Works | Sc Works | WSc Works | AI W |
|---|---|---|---|---|---|---|---|
| 1 | London School of Economics (LSE) | 2.04 | 3 | 5 | 1 | 1 | |
| 2 | World Bank Group | 3.25 | 2 | 1 | 2 | 4 | |
| 3 | National Bureau of Economic Research (NBER) | 3.77 | 20 | 26 | 4 | 2 | |
| 4 | International Monetary Fund (IMF) | 4.25 | 5 | 2 | 3 | 3 | |
| 5 | Department of Economics, Harvard University | 4.65 | 55 | 91 | 9 | 8 | |
| 6 | European Central Bank | 6.74 | 8 | 11 | 6 | 11 | |
| 7 | Department of Economics, University of California-Berkeley | 8.35 | 39 | 77 | 10 | 10 | |
| 8 | Economics Department, Massachusetts Institute of Technology (MIT) | 9.56 | 48 | 65 | 11 | 9 | |
| 9 | Department of Economics, University of Chicago | 9.81 | 37 | 61 | 8 | 6 | |
| 10 | Federal Reserve Board (Board of Governors of the Federal Reserve System) | 10.02 | 18 | 18 | 7 | 7 | |

- IDEAS ranking
  - Academic Oriented
  - Comprehensive ranking that include non-university institutions
- Full record linkage
  - Common word is removed (“university of”)
  - Take the highest score
  - Exactly 200 institutions were matched

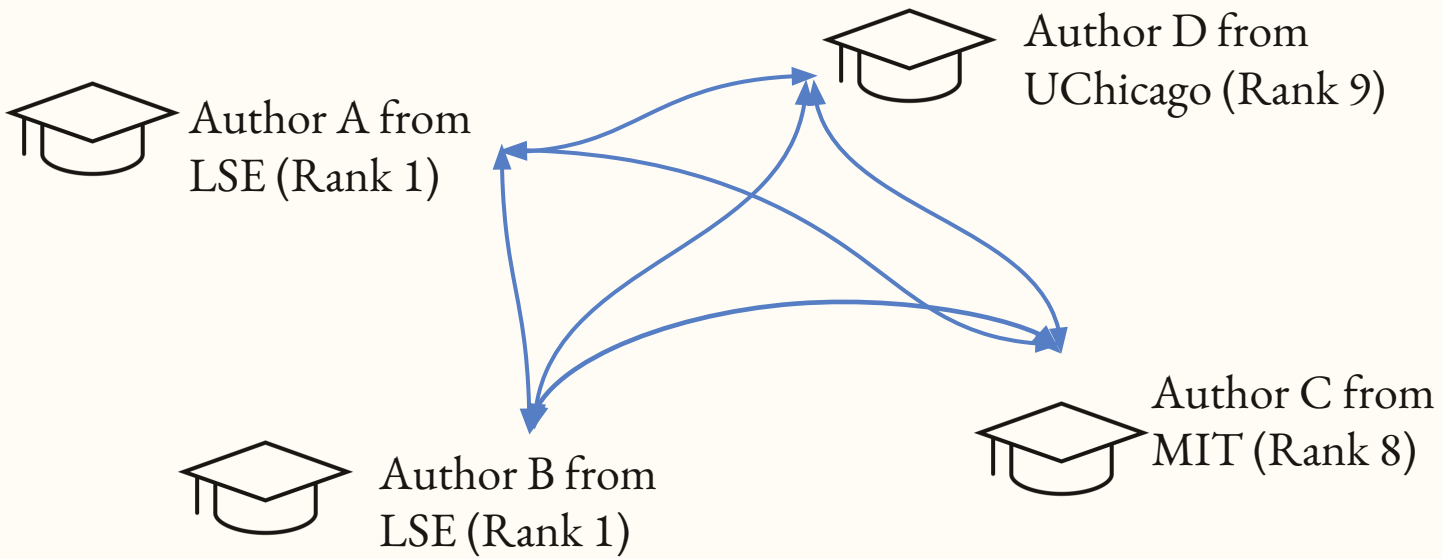
| affiliationid | name | email | match_... | univer_ins | Rank |
|---|---|---|---|---|---|
| au.aalto.fi | Aalto University | aalto.fi | 0.7962962... | Carleton University | 309 |
| kriot.kth.se | KTH Royal Institu... | kth.se | 0.7470084... | Beijing Institute of ... | 78 |
| uocla.ucla.edu | University of Calif... | ucla.edu | 0.682900... | Boston College | 100 |
| aei.aei.org | American Enterpr... | aei.org | 0.7599879... | Economic and Soci... | 212 |
| miot.mit.edu | Massachusetts In... | mit.edu | 1 | Massachusetts Inst... | 8 |
| ciot.caltech.edu | California Institut... | caltech.edu | 0.8677729... | Beijing Institute of ... | 78 |
| hu.harvard.edu | Harvard University | harvard.edu | 1 | Harvard University | 5 |
| su.stanford.edu | Stanford University | stanford.edu | 1 | Stanford University | 14 |
| uop.upenn.edu | University of Pen... | upenn.edu | 1 | University of Penns... | 55 |
| ai.amazon.com | Amazon Inc. | amazon.com | 0.6931818... | Chapman University | 448 |
| au.american.edu | American Univers... | american.e... | 0.746666... | Academia Romana | 129 |
| ac.amherst.edu | Amherst College | amherst.edu | 0.8002011... | Dartmouth College | 83 |
| cmu.cmu.edu | Carnegie Mellon ... | cmu.edu | 1 | Carnegie Mellon Un... | 360 |
| anu.anu.edu.au | Australian Nation... | anu.edu.au | 1 | Australian National ... | 99 |
| au.ariel.ac.il | Ariel University | ariel.ac.il | 0.7460317... | Drexel University | 196 |
| uoa.arizona.edu | University of Ariz... | arizona.edu | 0.7925925... | University of Alberta | 416 |
| asu.asu.edu | Arizona State Uni... | asu.edu | 1 | Arizona State Unive... | 177 |
| frboa.atl.frb.org | Federal Reserve ... | atl.frb.org | 1 | Federal Reserve Ba... | 257 |
| auob.aub.edu.lb | American Univers... | aub.edu.lb | 0.666379... | Central Bank of Irel... | 386 |

# Database Outlook



# Constructing Co-authorship links

Article – Author – Institution – Ranking

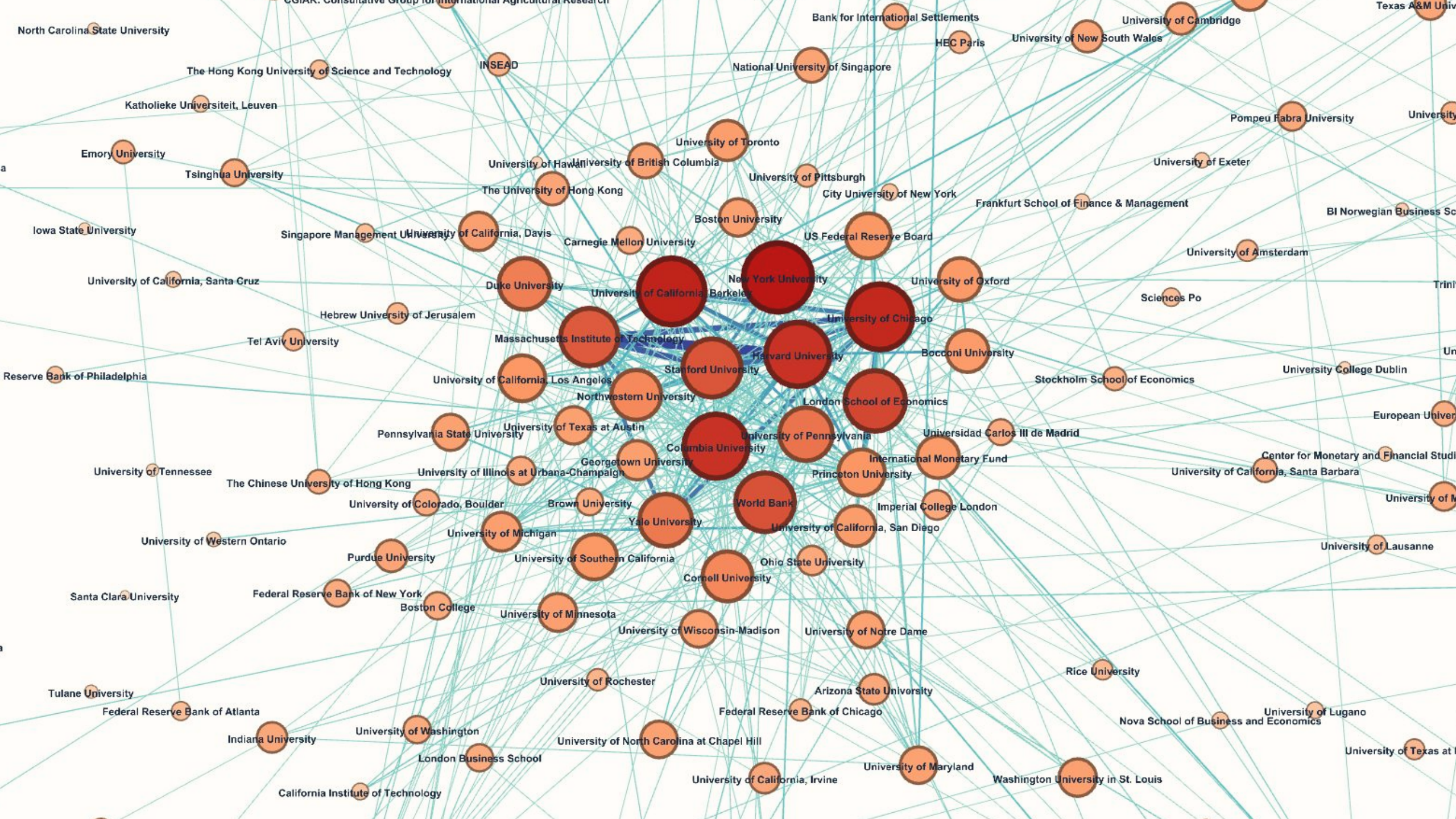


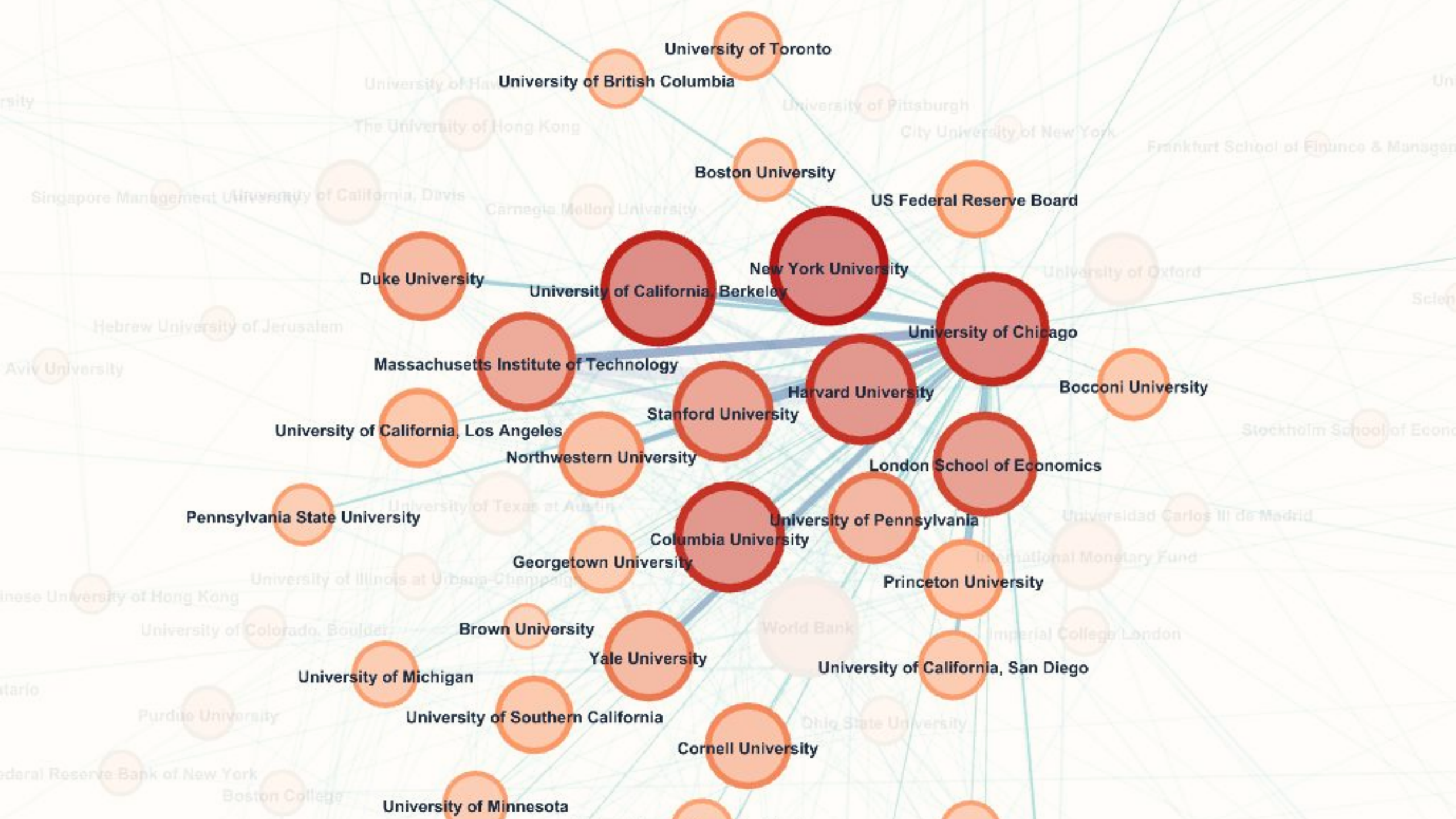

***LSE (Whole Dataset)***

| Name | Number of links |
|---|---|
| LSE | 10 |
| UChicago | 5 |
| MIT | 5 |

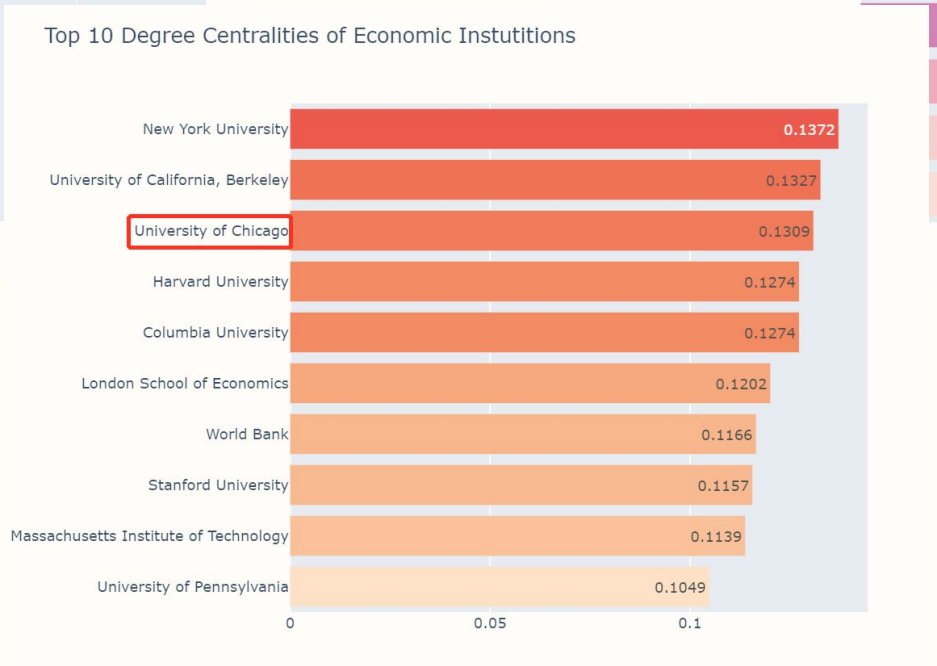
*(example)*

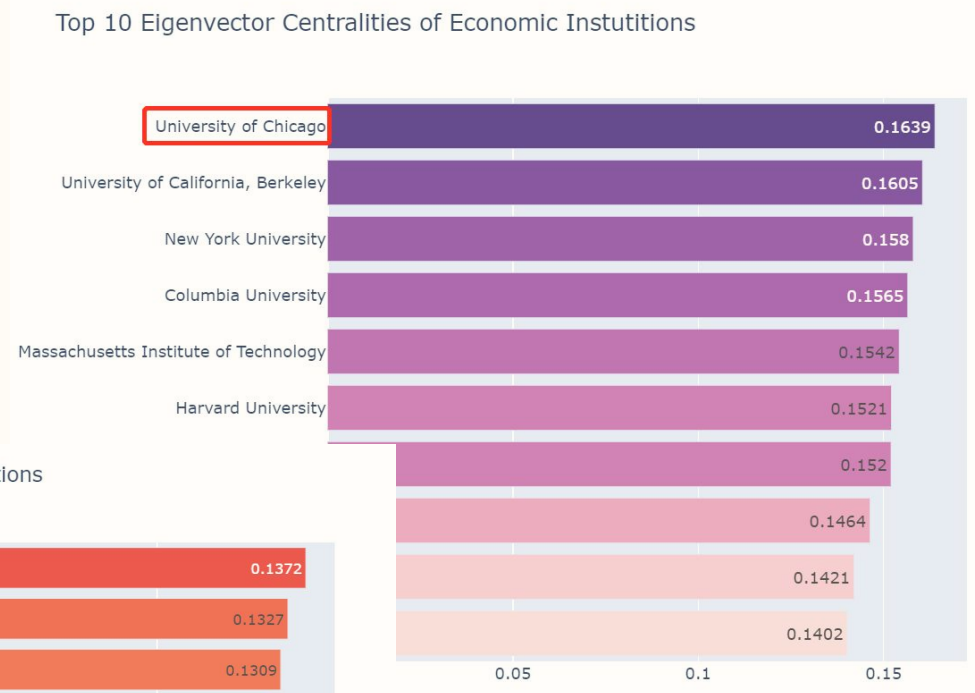
- Compile to Institution Level links


CGIAR Consultative Group for International Agricultural Research
North Carolina State University
Bank for International Settlements
University of Cambridge
HEC Paris
University of New South Wales
The Hong Kong University of Science and Technology
INSEAD
National University of Singapore
Katholieke Universiteit, Leuven
Pompeu Fabra University
Emory University
University of Toronto
University of Hawai'i
University of British Columbia
Tsinghua University
University of Pittsburgh
University of Exeter
The University of Hong Kong
City University of New York
Frankfurt School of Finance & Management
Boston University
Iowa State University
Singapore Management University
University of California, Davis
Carnegie Mellon University
US Federal Reserve Board
University of Amsterdam
University of California, Santa Cruz
Duke University
New York University
University of Oxford
University of California, Berkeley
Sciences Po
Hebrew University of Jerusalem
University of Chicago
Massachusetts Institute of Technology
Tel Aviv University
Harvard University
Bocconi University
Reserve Bank of Philadelphia
Stanford University
University College Dublin
University of California, Los Angeles
Stockholm School of Economics
Northwestern University
London School of Economics
University of Texas at Austin
Pennsylvania State University
University of Pennsylvania
Universidad Carlos III de Madrid
Columbia University
Georgetown University
International Monetary Fund
University of Tennessee
University of Illinois at Urbana-Champaign
Princeton University
Center for Monetary and Financial Studi
The Chinese University of Hong Kong
University of California, Santa Barbara
University of Colorado, Boulder
Brown University
World Bank
Imperial College London
Yale University
University of Michigan
University of California, San Diego
University of Western Ontario
University of Lausanne
Purdue University
University of Southern California
Ohio State University
Cornell University
Santa Clara University
Federal Reserve Bank of New York
Boston College
University of Minnesota
University of Wisconsin-Madison
University of Notre Dame
Rice University
University of Rochester
Tulane University
Arizona State University
Federal Reserve Bank of Atlanta
Federal Reserve Bank of Chicago
University of Lugano
Nova School of Business and Economics
University of Washington
Indiana University
University of North Carolina at Chapel Hill
London Business School
University of Maryland
University of California, Irvine
Washington University in St. Louis
California Institute of Technology


University of Toronto
University of British Columbia
Boston University
US Federal Reserve Board
Duke University
New York University
University of California, Berkeley
University of Chicago
Massachusetts Institute of Technology
Harvard University
Bocconi University
Stanford University
University of California, Los Angeles
Northwestern University
London School of Economics
Pennsylvania State University
University of Pennsylvania
Columbia University
Georgetown University
Princeton University
Brown University
Yale University
University of Michigan
University of California, San Diego
University of Southern California
Cornell University
University of Minnesota

# Centralities Analysis

# Quantifying Collaboration links

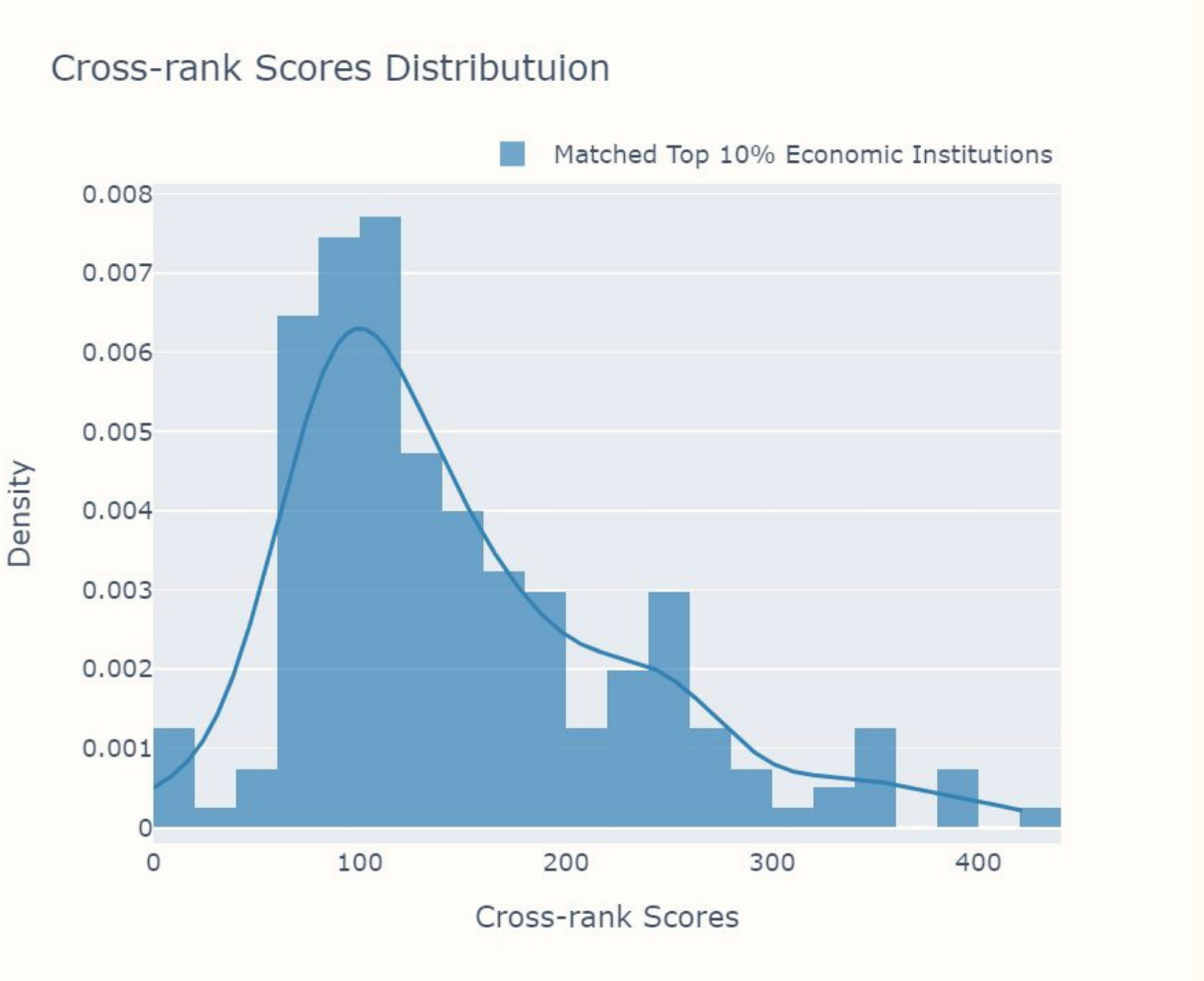
## (1) Institution level analysis

***LSE (Whole Dataset)***

| Name | Number of links | Contribution To the score |
|---|---|---|
| LSE (Rank 1) | 10 | 0 |
| UChicago (Rank 9) | 5 | (9-1) * 5 |
| MIT (Rank 8) | 5 | (8-1) * 5 |

*(example)*

Cross-rank score = (0 + (9-1) * 5 +(8-1) * 5) / 20 = 3.75



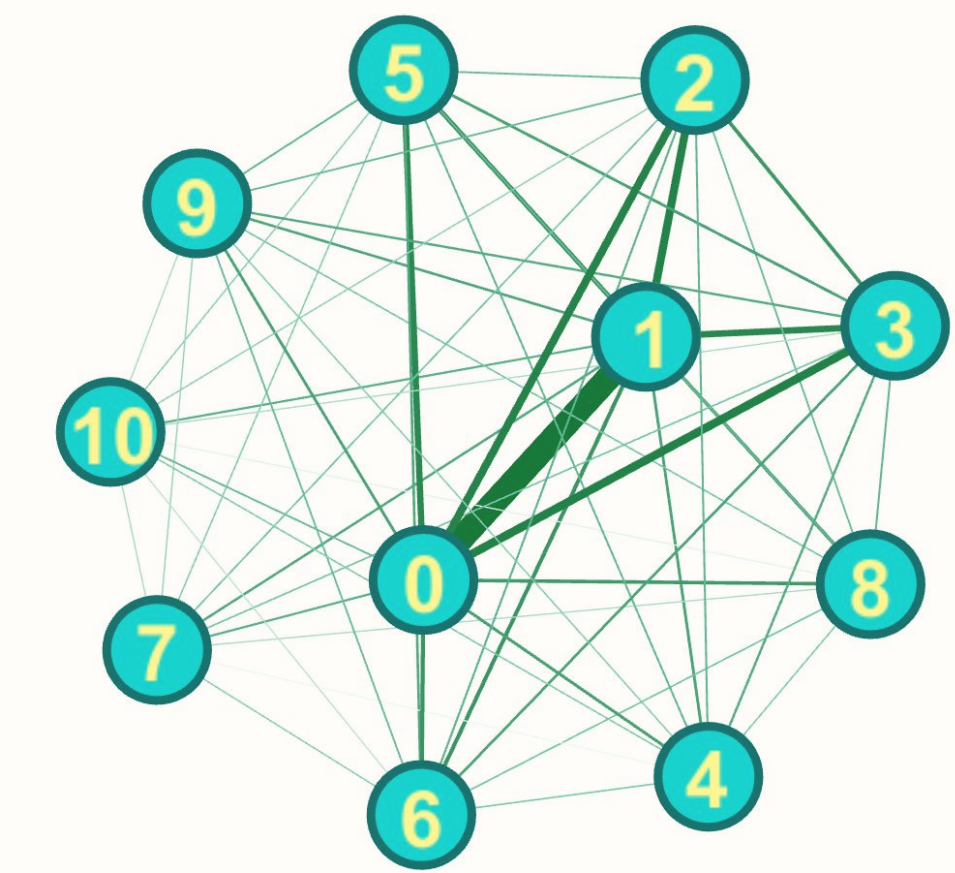
# Quantifying Collaboration links

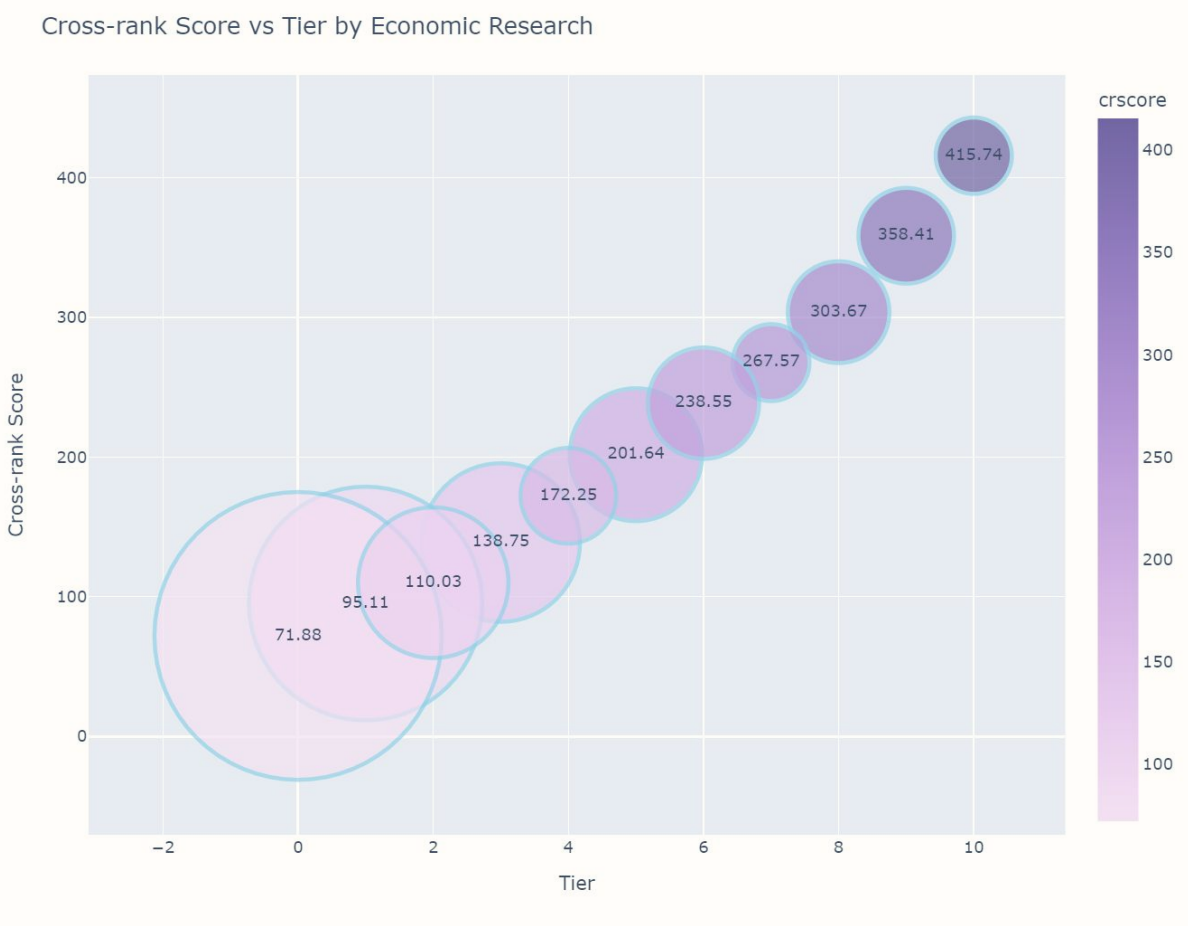
## (1) Institution level analysis

- Institutions that rank higher tend to collaborate more with each other - a low cross-rank score.
- Cross-ranking collaboration is more common for institutions that rank later.

# Quantifying Collaboration links

## (2) Institution Tier level analysis



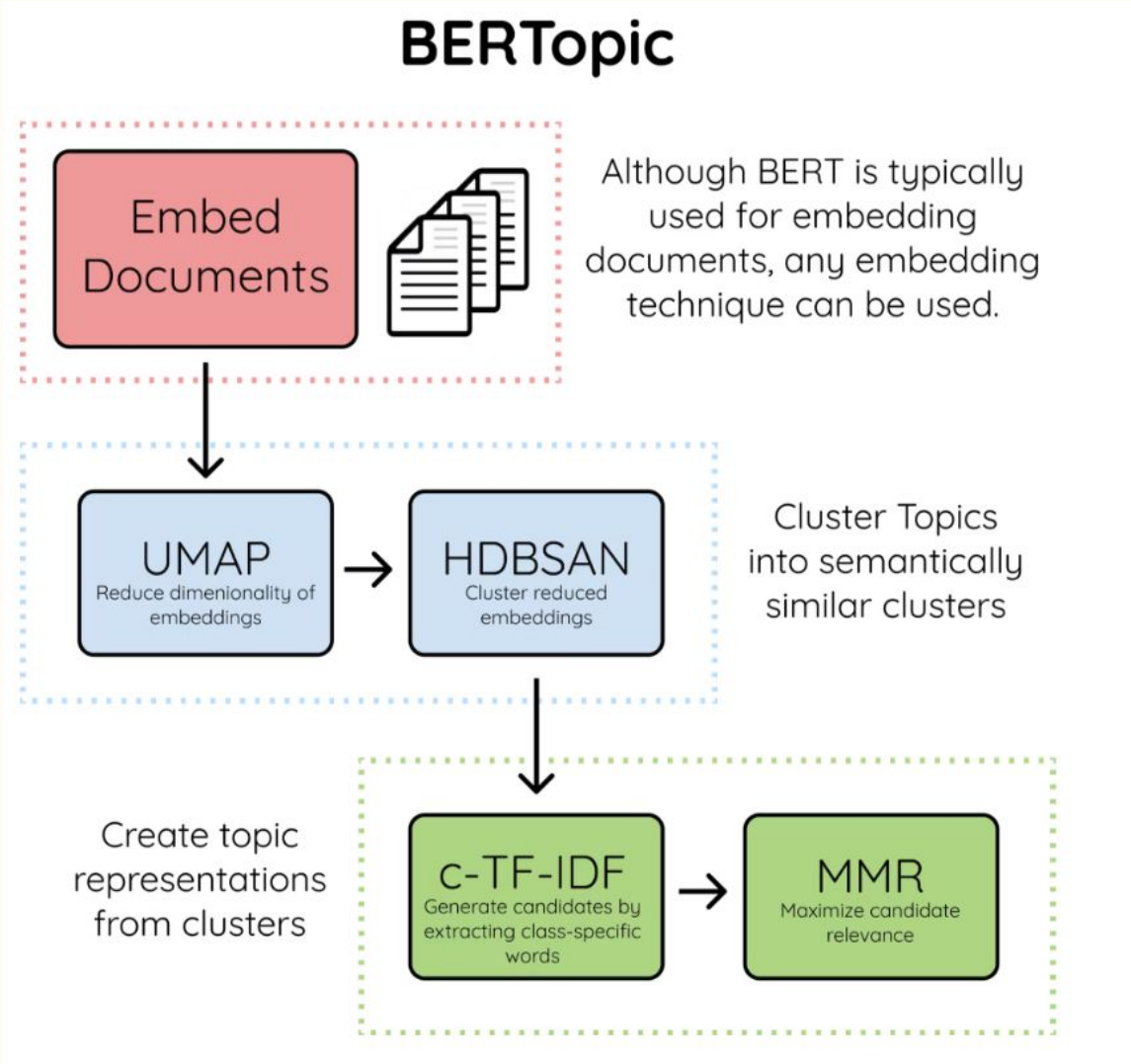
# Topic Modeling Analysis

## BERTopic Modeling

- An emerging TM technique that uses language embedding and class-based TF-IDF
- Original Paper (Grootendorst 2022): https://arxiv.org/abs/2203.05794
- Documentation: https://maartengr.github.io/BERTopic/index.html

## Advantages over LDA TM

- Easy to fit in model: only need to clear stopwords and re-concatenate into paragraphs
- Much more stable: results may vary a little across runs but not as much as LDA
- Much less training time: 15,000+ abstracts take less than 1 minute to train

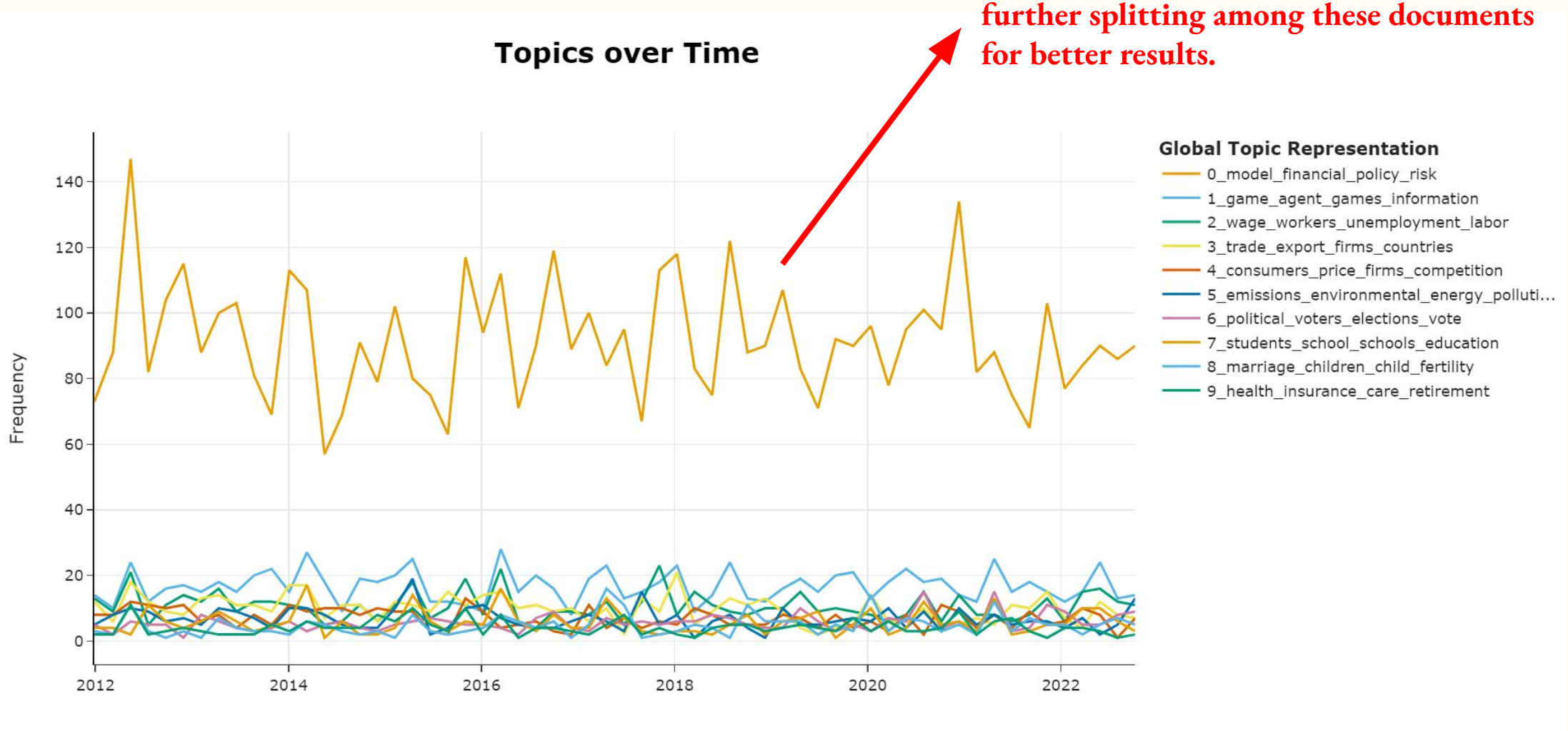



*Source:https://hackernoon.com/nlp-tutorial-topic-modeling-in-python-with-bertopic-372w35l9*

# Topic Modeling Analysis

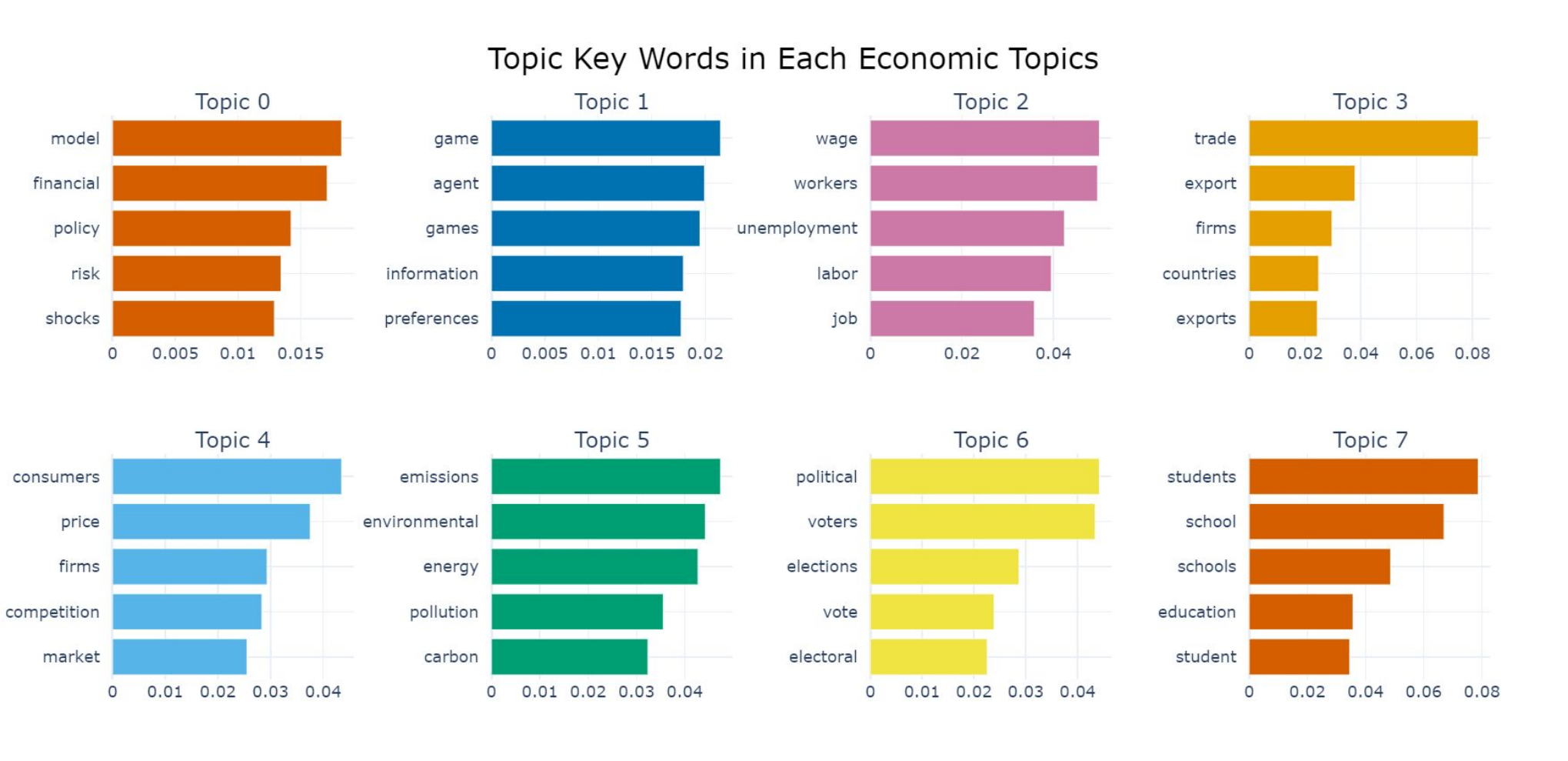
## Dynamic Economic Research TM over Time

**Dominating, but NON-trivial! Need further splitting among these documents for better results.**



# Topic Modeling Analysis

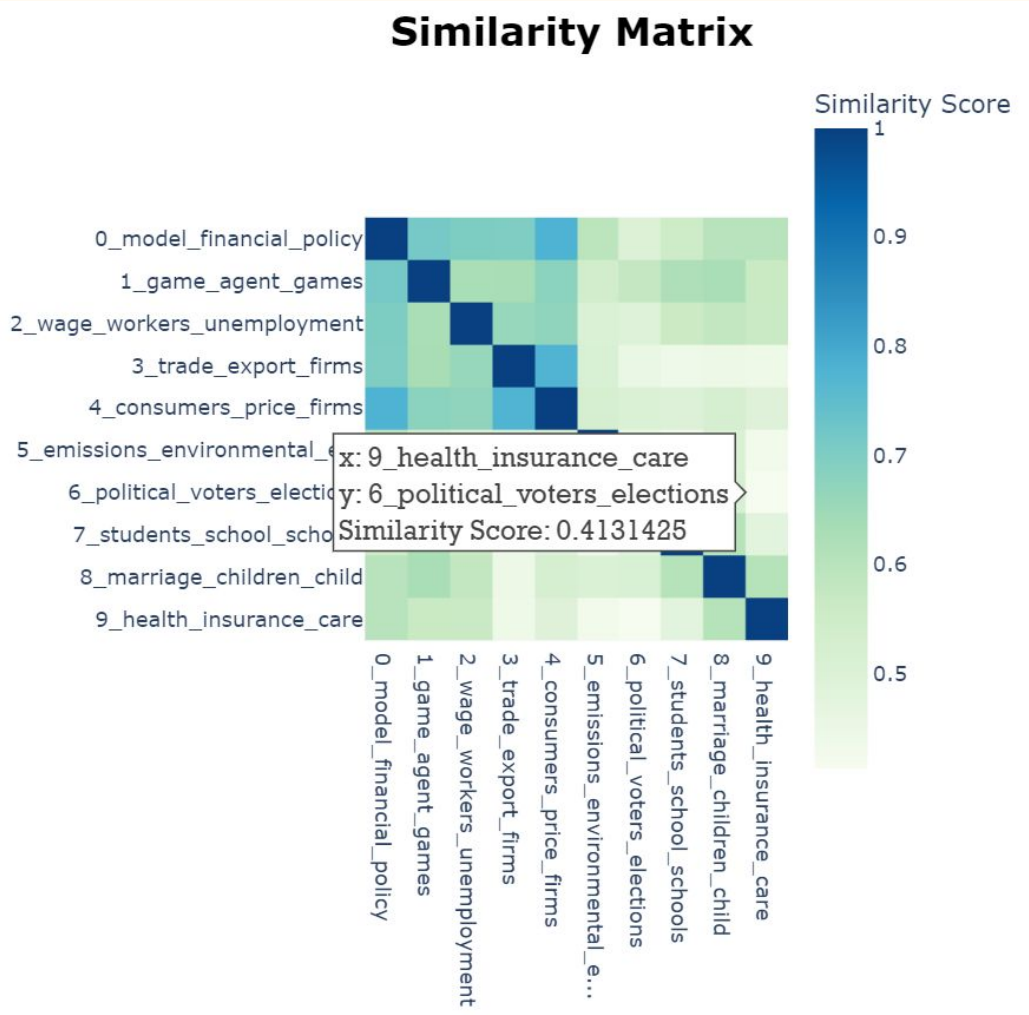
## Top 5 Keywords for each Topic:

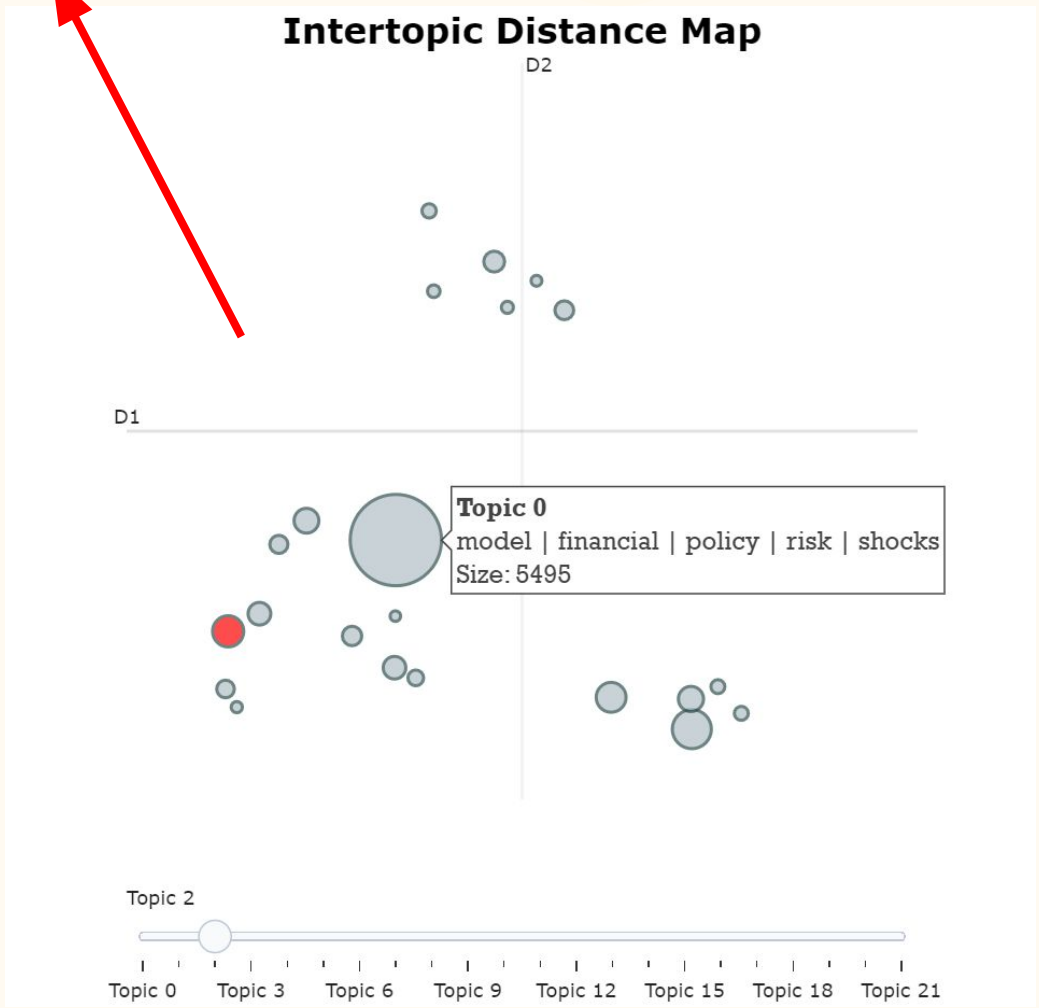


# Topic Modeling Analysis

## Similarity and Intertopic Distance

**All topics do NOT overlap on Intertopic Distance Map, so the training result should be fine!**



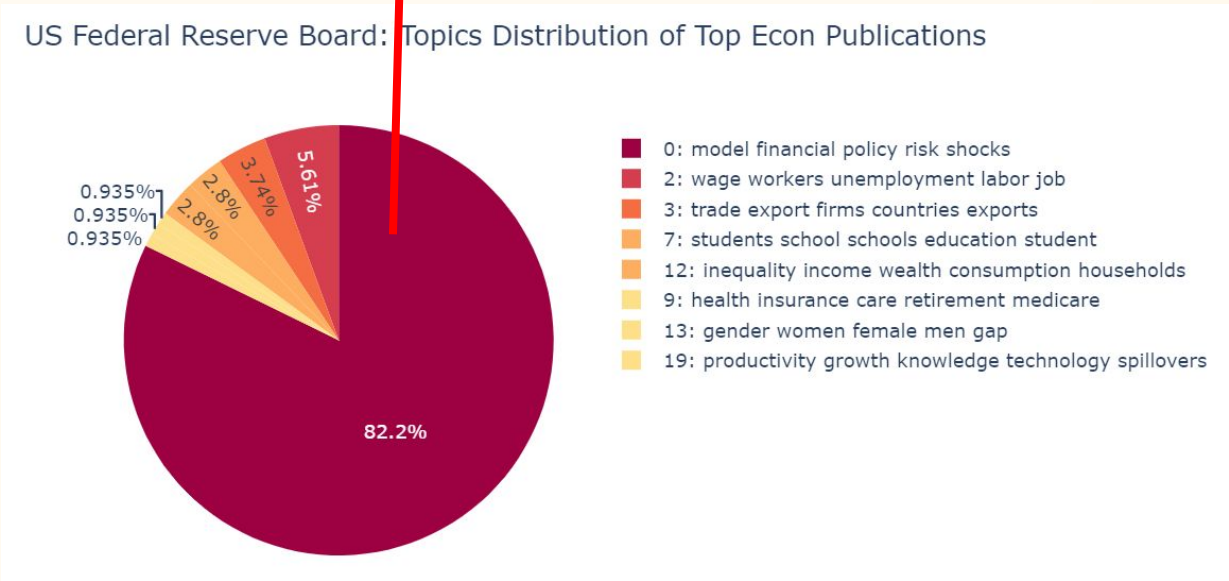



# Topic Modeling Analysis

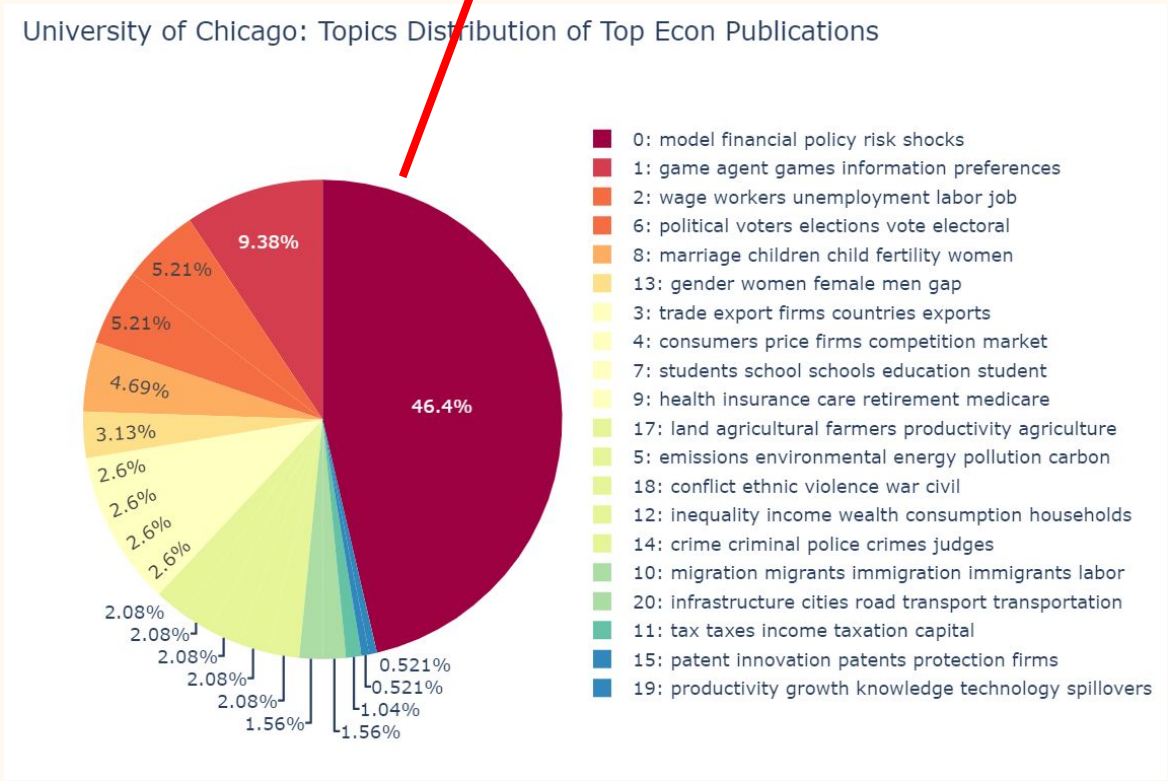
## Specialization Areas of Different Institutions

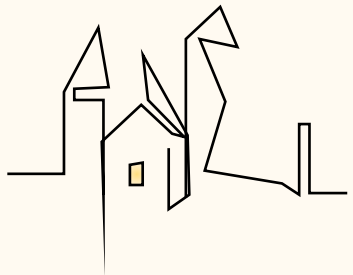
- Every document is assigned with a single MOST RELEVANT topic.
- We can use this result to see what topics an institution is focusing on.

**FED is mostly researching about "financial policy, risk, shocks", macro related topics. Make sense!**




**UChicago has a much broader interests of topics, almost covering all 21 topics we identify!**





# Take-home Points

## From Technical Experience:

- **SQL** is really convenient to structure datasets and extract relevant information!
- Except for speed, **Selenium** is way better than Requests for web scraping!
- **Plotly** is fantastic for fancier visualization and interactive plots!
- When doing Topic Modeling, try out **BERTopic** package! Do NOT stick to Gensim LDA.

## Social Science Significance:

- We form a whole picture of the current Economic academia by 1): identifying its collaboration patterns on Top Econ Journals through network analysis 2): research topics evolvement across time and institutions through dynamic topic modeling.
- ***From network***, Top tier institutions are more likely to collaborate with institutions of similar levels; but for lower tier institutions, they often collaborate with those from very different tiers. In other words, an institution's cross-rank/tier collaboration likelihood decreases as its tier/rank gets higher.
- ***From topic modeling***, we see topics are stable in trends across time with some fluctuations but can vary a lot across different institutions.
- ***Future work***: 1) explore more parameters in network like centrality, shortest path; 2) further split the dominating topic into subtopics for better classification; 3) combing network and TM, see collaboration patterns under different topics.

This is the end of Presentation.
Thanks for Listening!

# Q & A