Visualizing Institutional Scholar Relationships (VISR)

Benjamin Vargas, Gustav Rödström, Karan Kishinani, Peter McAughan, William Broniec, Yang Hu

Problem

Many relationships between academic and industrial institutions are clear and explicit through established programs and funding; many of these relationships are also informal and can't be easily identified.

Why?

- Citation graphs in the past have been focused on granularity and direct connections between research groups and individuals
- This type of analysis ignores the broader understanding of the relationships

Our Solution



graph analytics clustering algorithms, we want to and visualize capture institutional relationships so that they can be better understood and recognized.

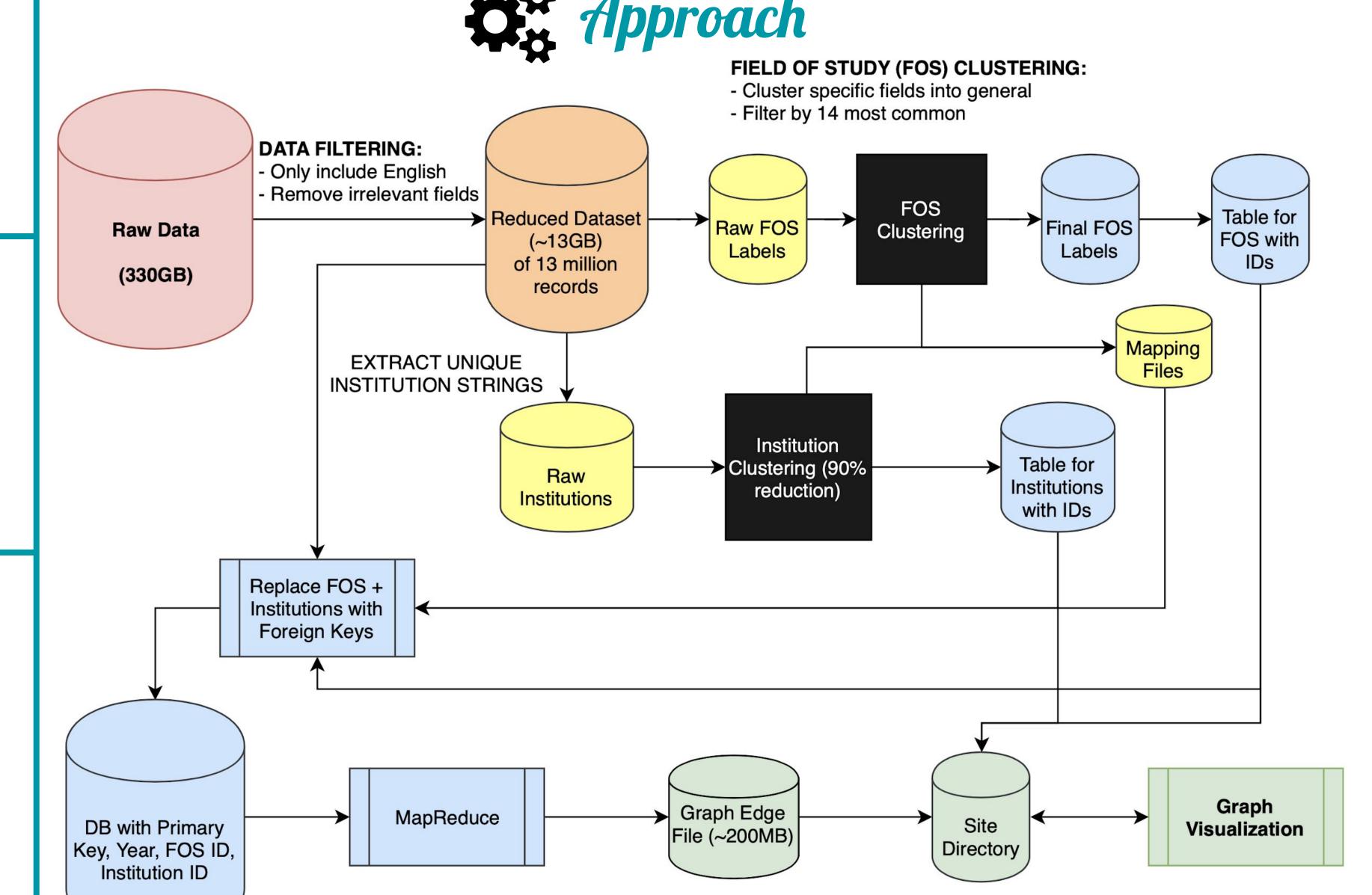
The Dataset

- Microsoft Academic Graph (MAG) by AMiner
- Originally 166,192,182 records and >300 GB!
- Includes publication author(s), institutions, etc.



What's New?

- Focus on *Institutions* instead of individual researcher connections
- Analyze the data using a more holistic perspective to identify unexplored trends



Reduction + Clustering String i continue co<u>ntinue</u> String i + 1 String i + 1 remove @#\$%, trailing and prepending spaces String i fingerprint Group label: X split by space Words bank (corpus) Bag of words convert to convert to Tf-idf vector Tf-idf vector compute if score < if score >= similarity threshold threshold String i Group label: X + 1 Group label: X classified as new group **X** + **1** Words bank (corpus) Words bank (corpus)

- Map specific fields of study & institution names into broad categories based on similarity
- Achieve fuzzy string matching using Tf-idf and fingerprinting
- **Problem**: Conventional Tf-idf is not feasible due to O(n²) runtime for multiple clusters

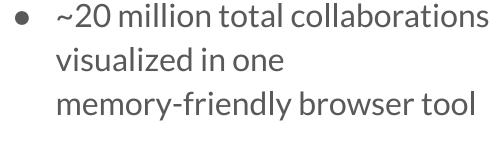
current group X

- Assumption: After sorting, similar institutions may appear closer, allowing linear runtime computation
- Solution: Reinitialize word bank when new cluster is created

Visualization Implemented using Field: SOCIAL SCIENCE Search for an Institution Search Load EASTERN MEDITERRANEAN UNIVERSITY **OHIO STATE UNIVERSITY** UNIVERSITY OF NORTH CAROLINA INSTITUTE FOR HEALTHCARE POLICY PENNSYLVANIA STATE UNIVERSITY **COLUMBIA UNIVERSITY UNIVERSITY OF MICHIGAN** F PIRAEUS UNIVERSITY OF MINNESOTA STANFORD UNIVERSITY INDIANA UNIVERSITY **UNIVERSITY OF ARIZONA DUKE UNIVERSITY** RUTGERS UNIVERSITY UNIVERSITY OF PENNSYLVANIA MASSACHUSETTS INSTITUTE OF TECHNOLOGY UNIVERSITY OF CALIFORNIA CALIFORNIA STATE UNIVERSITY UNIVERSITY OF BRITISH COLUME YALE UNIVERSITY UNIVERSITY OF WISCONSIN CARNEGIE MELLON UNIVERSITY **UNIVERSITY O F PENNSYLVANIA TEMPLE UNIVERSITY** NORTHWESTERN UNIVERSITY IAL UNIVERSITY PRINCETON UNIVERSITY UNIVERSITY OF CHICAGO UNIVERSITY OF SOUTHERN CALIFORNIA **NEW YORK UNIVERSITY** Field of Study Filtering

Results & User Feedback

- 330 GB of Academic Paper data cleaned and condensed into 8.5 million edges and 1 million nodes
- Average node degree (collaboration): 4.1 collaborations
- 47,000 Paper topics clustered into 14 overarching fields

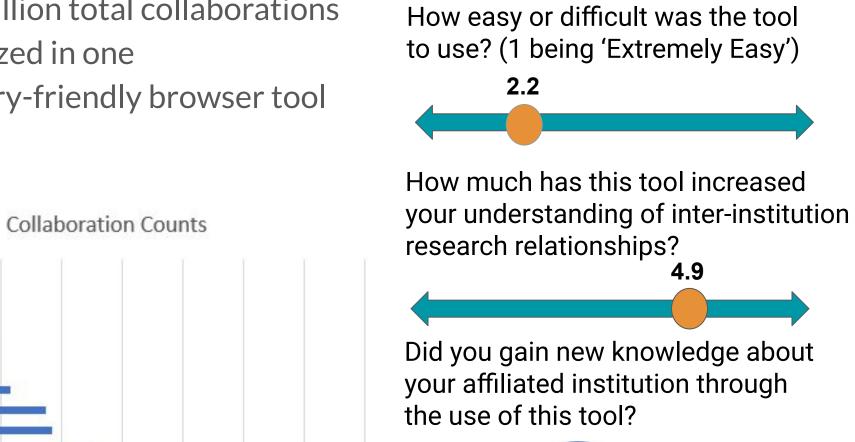


500000 1000000 1500000 2000000 2500000 3000000 3500000

Computer Science

Engineering

Medicine



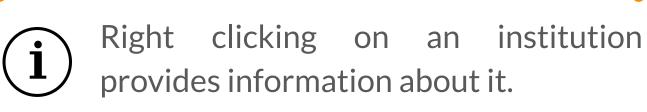
77.8%

Yes

No

N/A

Enables filtering by field of study and search for a specific institution using autocomplete



Functionality

Ability to scroll around the graph, pin

and highlight nodes by double clicking

Emphasizes edges with high weight by

color and nodes with large degrees by

