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

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

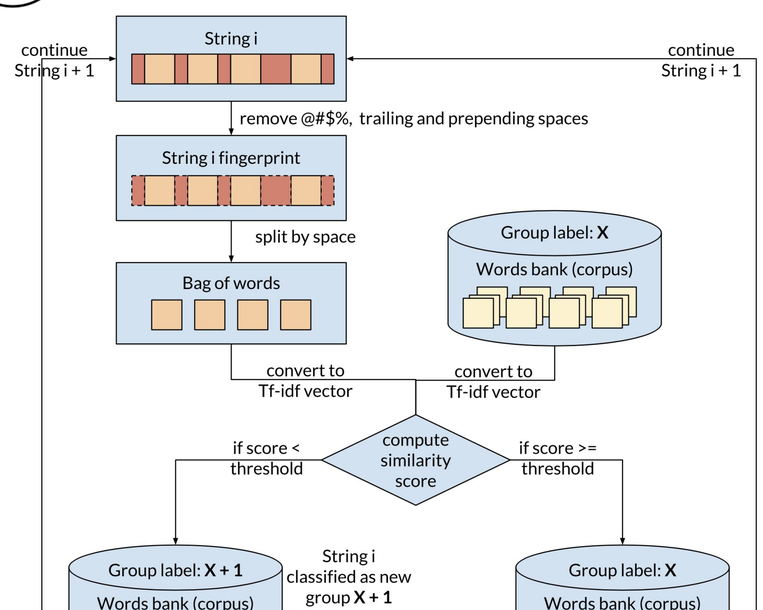
The Dataset

- 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

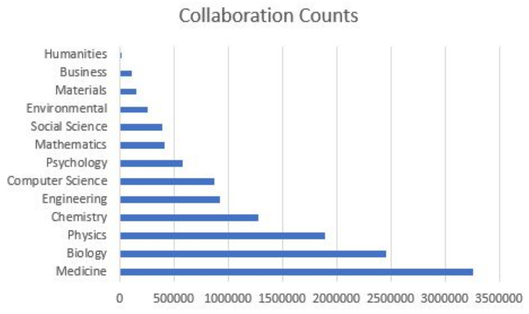
Reduction + Clustering



- 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 $O(n^2)$ runtime for multiple clusters
- ***Assumption***: After sorting, similar institutions may appear closer, allowing linear runtime computation
- ***Solution***: Reinitialize word bank when new cluster is created

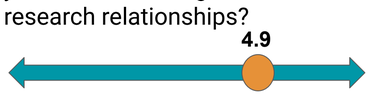
Results & User Feedback

- 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
- ~20 million total collaborations visualized in one memory-friendly browser tool

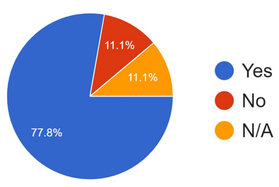


How easy or difficult was the tool to use? (1 being 'Extremely Easy')

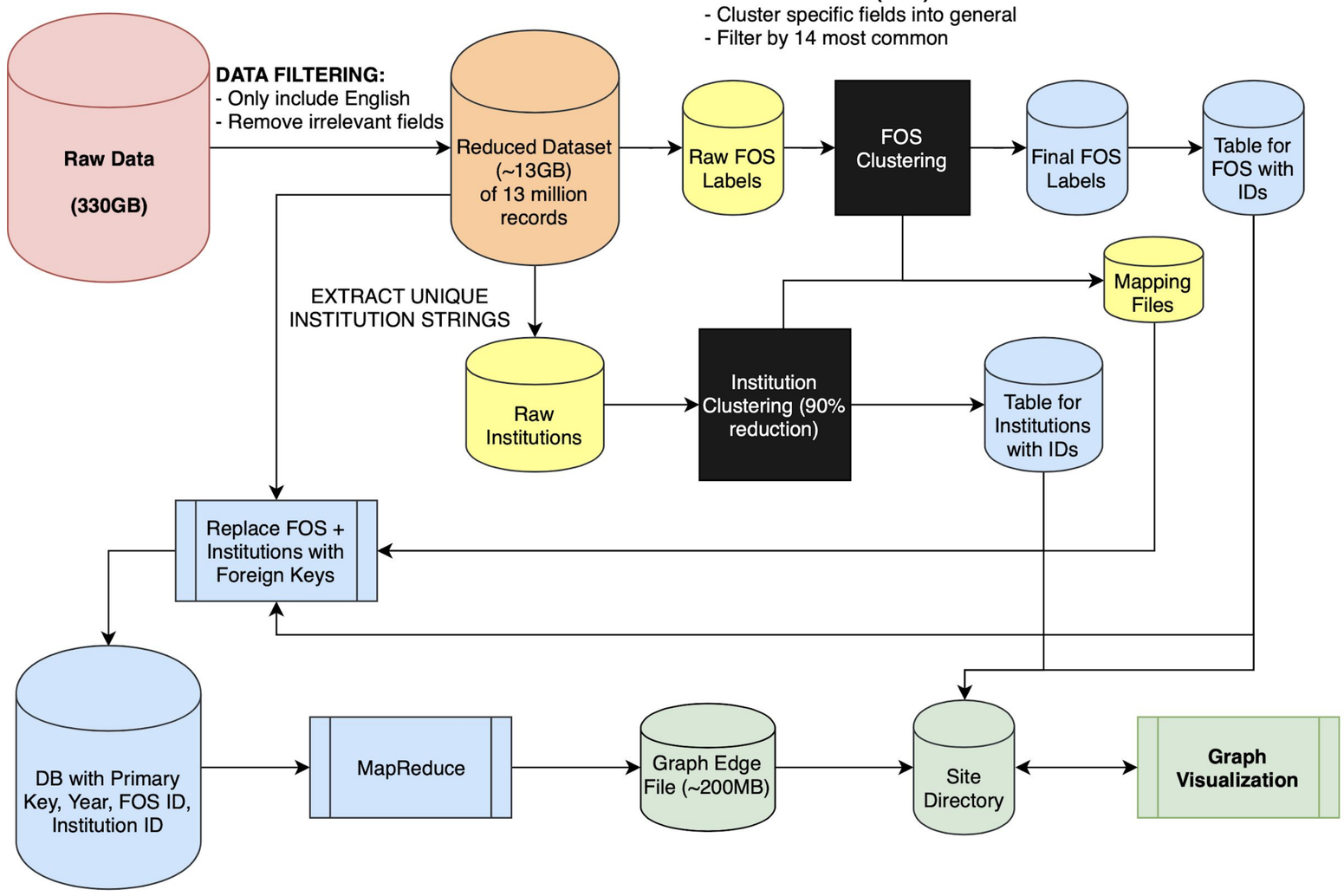
How much has this tool increased your understanding of inter-institution research relationships?



Did you gain new knowledge about your affiliated institution through the use of this tool?

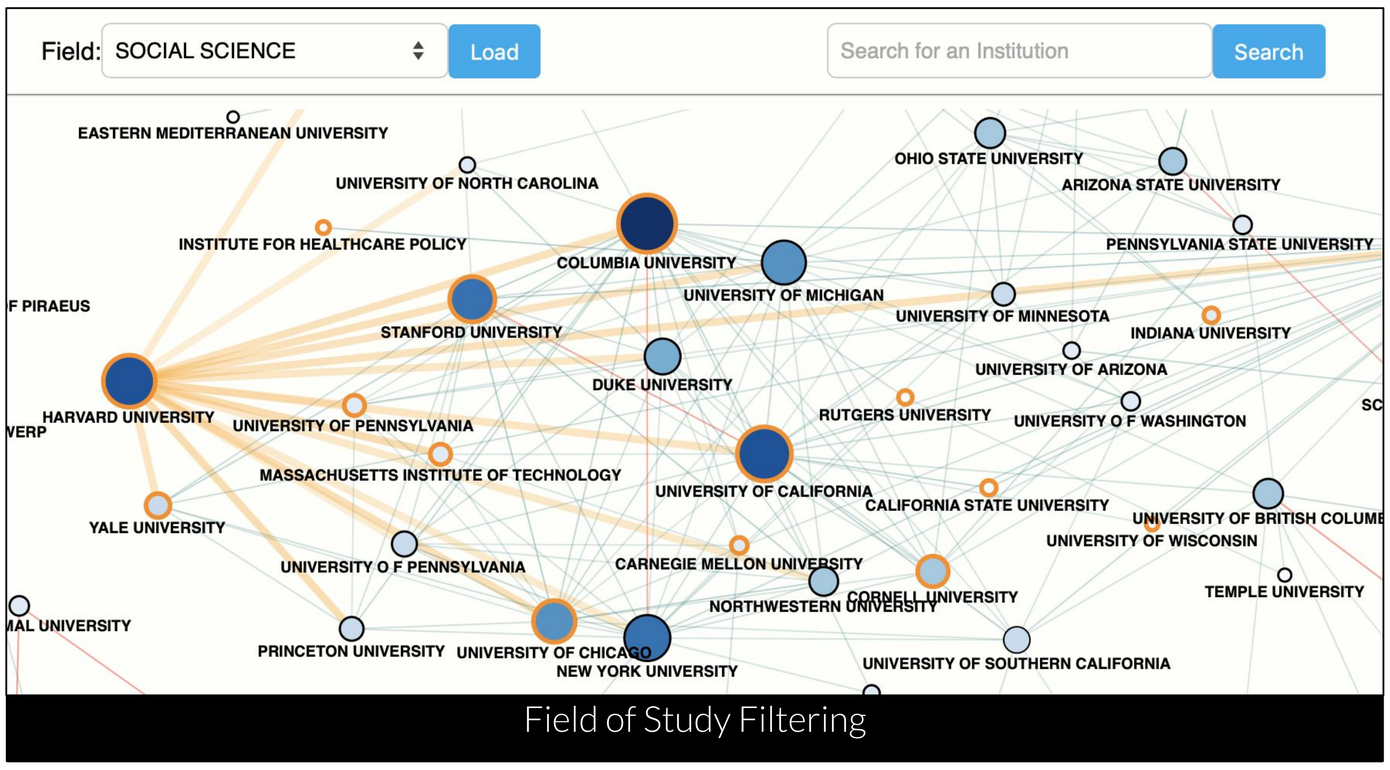


Approach



Visualization

Implemented using D3



Field of Study Filtering

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 size

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

Right clicking on an institution provides information about it.

Search by Institution