# Jaks

**Github Repository**

## Team


| | | |
|---|---|---|
| Maks Cegielski-Johnson | maks.cegielski@gmail.com | u0836524 |
| Jake Pitkin | jakepitkin@gmail.com | u0891770 |
| Jackson Stafford | jackson.stafford@utah.edu | u0342191 |


## Background and Motivation

Finding research papers online is a relatively easy process currently but comparing them is not. We want to make an easy way to visualize research papers and their relationships to each other. A user will be able to make a customized query based on a combination of researcher's names, affiliations, topics. The goal is to develop a useful way to discover new papers that could be relevant to your area of research you otherwise would have not been exposed to.

## Project Objectives

The space of all research papers is vast (we found a dataset of over a million papers) and we will have to be creative to visualize the information.

User's typically will have a concentrated interest when exploring research papers such as: a researcher's name, a specific university or a research area. Additionally a user might want to see the collaboration between a group of researchers.

Make filtering and exploring a vast collection of research papers an initiative process that is accessible to all researchers.

Learn how to create large graph visualizations in D3.

Mining and cleaning a medium size real world dataset. Working with the limitations of real world data.

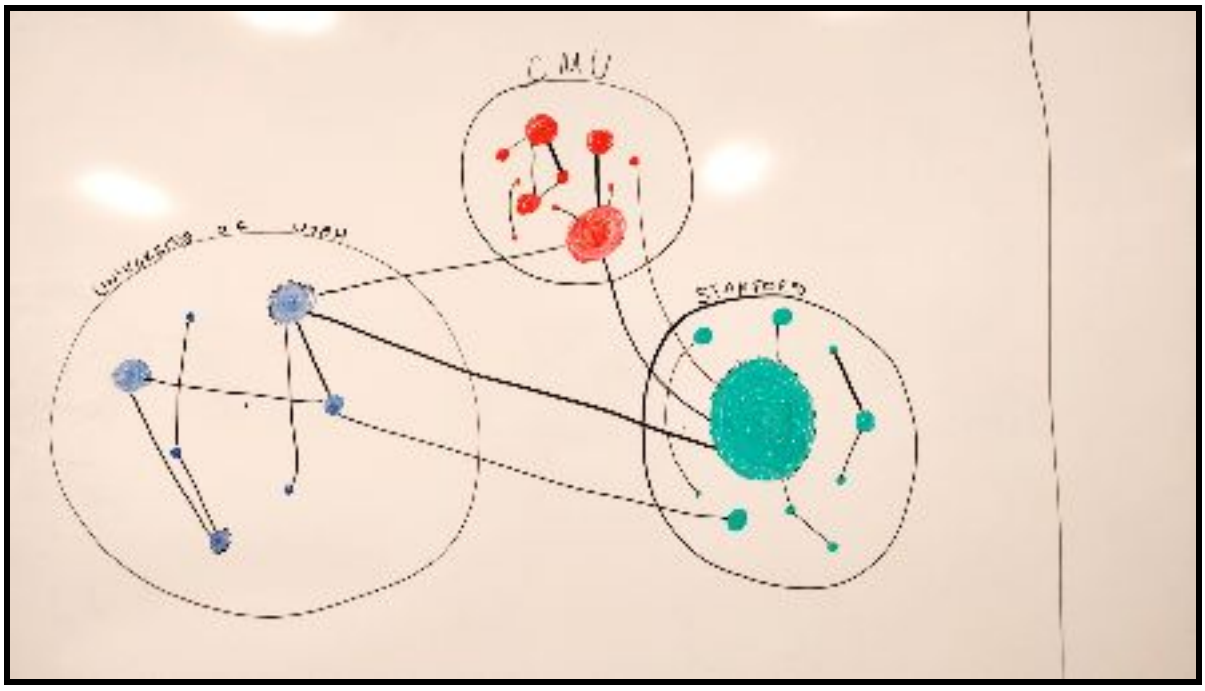
## Data

We will be using arxiv.org's paper archive as data for our visualization. Using their bulk data tool (https://arxiv.org/help/bulk_data) we will acquire a collection of research papers.

## Data Processing

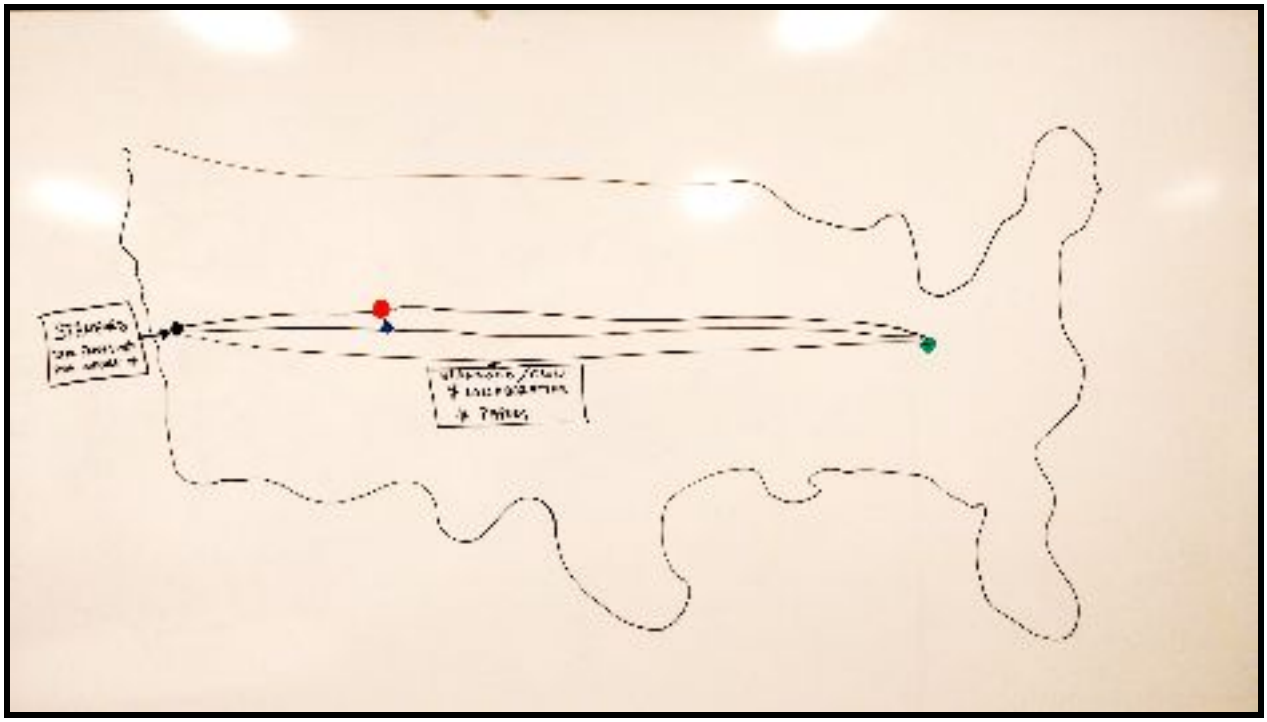
Data is in xml format, so different sections are easily accessible. Data preprocessing may likely be limited to aggregating author names and institution cleanup; e.g. F. Li vs Feifei Li.

# First Design



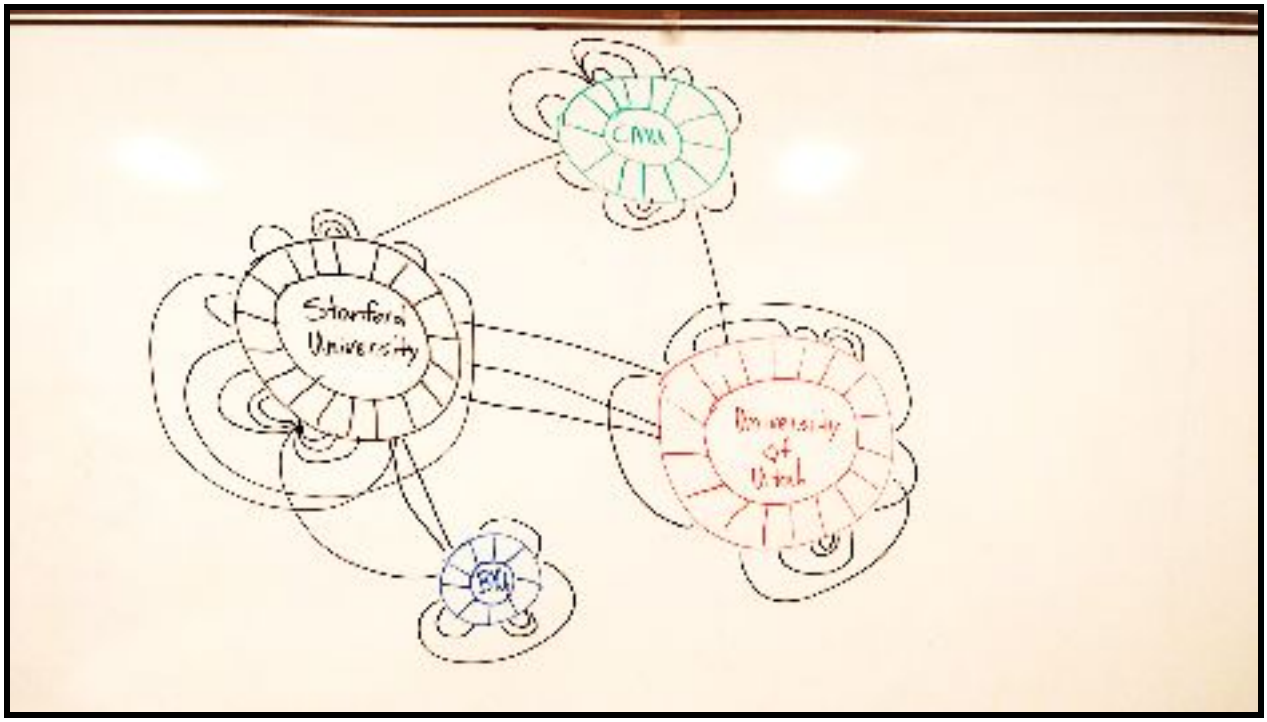

Our first attempt at a design is an undirected graph. Nodes represent researchers and edges represent papers. Nodes are colored based on university affiliation and are spatially near each other. The size of a node represents the numbers of papers a researcher has produced.

## Second Design



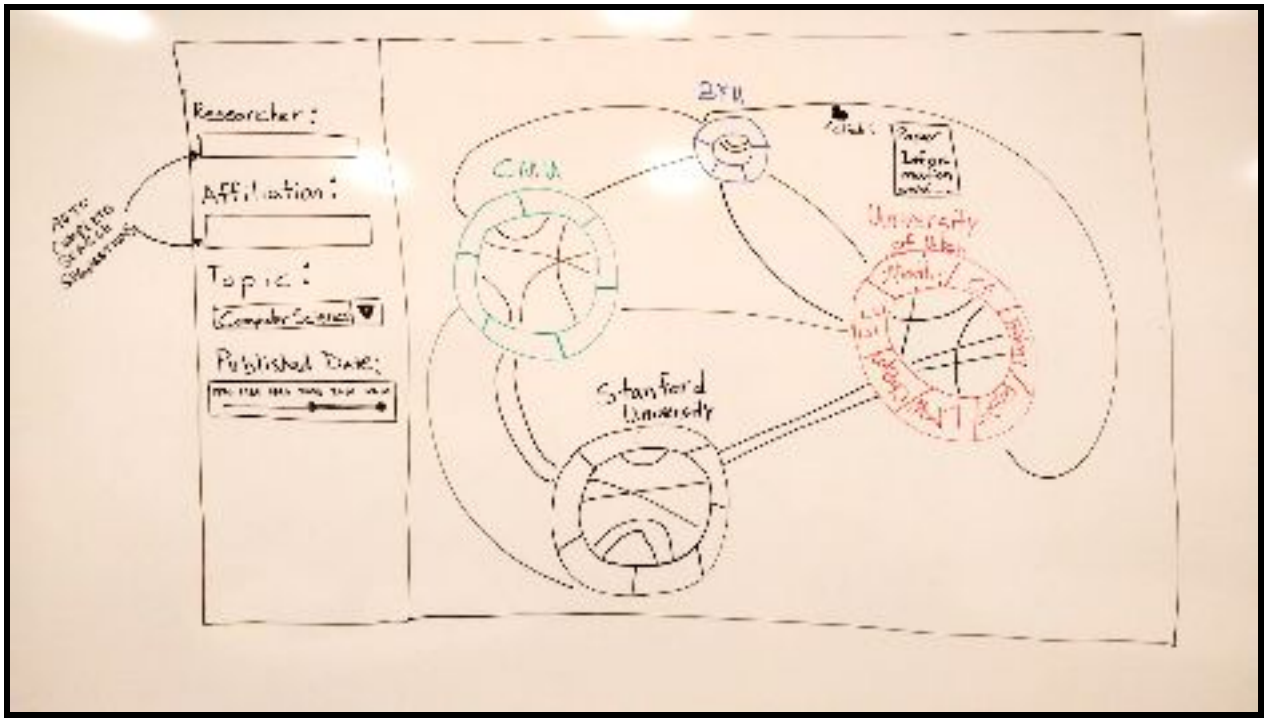
Our next idea was to use a geographic representation of the United States to place each of the universities. Each node represents a university and each edge is an aggregate of collaborations between universities. Clicking a university node will display information about the researchers that work at the given university and their internal collaborations. Clicking an edge will display a collection of papers that were produced by a collaboration between researchers at the two respective universities. A limitation of this design is too much information will exist in the aggregations.

# Third Design




Our third design is the design that will serve as our prototype for our project. Each university is represented as a ring graph where each section in the ring is a researcher. Each edge in the graph will be a paper between two researchers. We improved on this design and go into more detail on the next page.

# Final Design




This design will combine all the ideas of the previous three designs. Each university will have a ring where each sector is a different field. Between each sector will be “flow” lines that signal inter-disciplinary collaboration on papers, aggregated between two fields. Clicking on a sector will expand that sector into authors by that field, with links between each other, as well as aggregated links with other disciplines in the university. Meanwhile, there will also be flow lines between universities and their disciplines/authors, signalling joint papers between different schools. Clicking on a professor or their link will highlight all links from that professor, showing who they have worked with.

The university rings will be placed around the SVG based roughly on their geolocation relative to other school, similarly as we did with states in homework 6. Clicking on edges when the context is at author level will provide link(s) to the PDFs of papers by that author and the link.

There will also be a menu for filter, allowing for searching for different authors, filtering out schools and topics, as well as selecting a range of years of publications to display.

## Must Have Features

- Display the initial network of researchers, grouped by university affiliations.
- Filtering the visualized nodes and edges by the name of a researcher.
- Filtering by the name of a university.
- Filtering by research area.
- Display a list of paper links when hovering over an edge.
- Search by professor name

## Optional Features

- Autocomplete in search fields.
- Visualize researcher's statistics with points of interest.
- Filter the visualization based on a given time range.
- Aggregate links between different universities based on different properties
- Place Universities roughly by geolocation (like the states in HW 6

## Project Schedule

| Class Week | Member | Responsibility |
|---|---|---|
| Week 10 | Jake | Initial research and design of code architecture for the project. |
| Week 10 | Jack | Mine and clean the data. |
| Week 10 | Maks | Mine and clean the data. |
| October 31st: Required lecture for peer-to-peer feedback | | |
| Week 11 | Jake | Start to implement an initial prototype of the visualization. |
| Week 11 | Jack | Start to implement an initial prototype of the visualization. |
| Week 11 | Maks | Back-end functions and model such as filtering. |
| Week 12 | Jake | Paper tooltips. |
| Week 12 | Jack | Positioning of universities. |
| Week 12 | Maks | Work on the UI. |

| | | |
|---|---|---|
| Week 13 | Jake | Get feedback on project and make suggested improvements. |
| Week 13 | Jack | Help with the suggested improvements. |
| Week 13 | Maks | Help with suggested improvements. |
| Week 14 | Jake | Final bug fixes, polish, and prepare for presentation. |
| Week 14 | Jack | Final bug fixes, polish, and prepare for presentation. |
| Week 14 | Maks | Final bug fixes, polish, and prepare for presentation. |