

**CS6630 Project
Visualizing CS Collaborations**

Process Book

Youjia Zhou, Wenzheng Tao, Mingxuan Han

11/09/2018

Contents

1 Member Information	1
2 GitHub Repository	1
3 Overview and Motivation	1
3.1 World View	1
3.2 Contrast view	2
3.3 Certain university view	2
4 Data Processing	2
4.1 Dataset	2
4.2 Data Processing	2
5 Visualization Design	2
5.1 Overview	2
5.2 Sketches	3
5.3 Prototype	4
5.4 Implementation	5
6 Future Plans	7

1 Member Information

Member 1: Youjia Zhou, zhouyj96180@gmail.com, u1208920

Member 2: Wenzheng Tao, wenzheng.tao@utah.edu, u1210098

Member 3: Mingxuan Han, u1209601@umail.utah.edu, u1209601

2 GitHub Repository

<https://github.com/zhou325/dataviscourse-pr-VisCsCollaborations>

3 Overview and Motivation

The DBLP Computer Science Bibliography dataset contains more than 1.2 million bibliographic records. Hence, for researchers, it is a useful tool to trace the academic works and to get bibliographic details when composing the list of references for the new papers.

While there are a lot of works on cs rankings based on DBLP dataset, we found that it is also interesting to study the collaborations in computer science between universities using this dataset. The motivation of this project is from a small talk. One of our group member would like to know which university has the most relationships with other institutions in computer science, indicating the willingness of this university to communicate with other institutions over the world. **Based on the DBLP dataset, we define such relationship as the number of publications two institutions have worked with together.** And we believe that we are not only people who are interested in such kind of questions, like which university are the most active in computer science and which institution they should go if they are interested in doing research in certain specific areas in computer science. It is become a natural choice for us to run this project.

Overall, our project is supposed to provide the interactive visualization of worldwide affiliations research achievements in top CS conferences. Publication amount as well as research cooperation will be displayed together. We will use group of charts to show the relationships among the universities in the dimensions of location, topic field, and time. According to the data scales, the charts can be divided into three groups: world view, contrast view and certain university view.

3.1 World View

Firstly, we put every university on a world map so that it will be convenient to watch and discover the geographical distribution of research affiliations. Users can select the period and universities of interest, and the corresponding data will be displayed. We link universities that have collaborated in CS publications so that users will be able to see the development of research affiliations among selected universities and how this relationship changes over time. Meanwhile, users can recognize how many cooperative companies one university owns and among its collaborators who is the closest one. In addition, we also aims to visualize the overview correlations between the CS rankings, closeness geometrically and cooperation variety in this part.

3.2 Contrast view

Our visualization will help users look at the differences between selected universities in CS research achievements, mainly through the contrast-view group of charts. It allows the user select the universities by clicking on the nodes or other methods, such as drawing a rectangle to include the nodes, which is helpful to build up a contrast among universities in a certain area. We will apply bar chart to compare the total publication amount in four sub fields of CS, those are AI, System, Theory and Interdisciplinary Areas. And there will be a parallel ranking table, where the user could have a precious overview of what are the ranks of the universities in total and the subfields. Also, the user could see and resort a table through clicking on different domain to see the exact ranking of the universities.

3.3 Certain university view

If the user is interested in learning more information of a certain university, our visualization allows the user selecting it and see the information such as university name, CS ranking, fields of research, number of cooperative universities and geographical parameters in a note panel. By click the certain field, the user will be able to see the ratio of the corresponding conferences in a donut chart.

4 Data Processing

4.1 Dataset

There are two datasets we need in this project. The first one is DBLP dataset and the second one is geographical dataset of each institution listed in the DBLP database. The second dataset would include two parts, first part is latitude and longitude of those institution in the world map and the second part is the name, number of publications and color data of each institution.

The original DBLP data is from <https://dblp.uni-trier.de/xml/>. However, the raw dataset is very large and difficult to clean. Fortunately, we found a cleaned dataset which contained all information we need from <https://github.com/emeryberger/CSRankings>. It is a summary of information for each article, which combines information of all authors.

4.2 Data Processing

With some transformations, we got two datasets from the cleaned data. "collaborations.json" is an overview of collaborations between institutions with each element representing an institution and its collaborating relationships with other institutions. And "collaborationsDetails.json" contains more detailed information of conferences.

5 Visualization Design

5.1 Overview

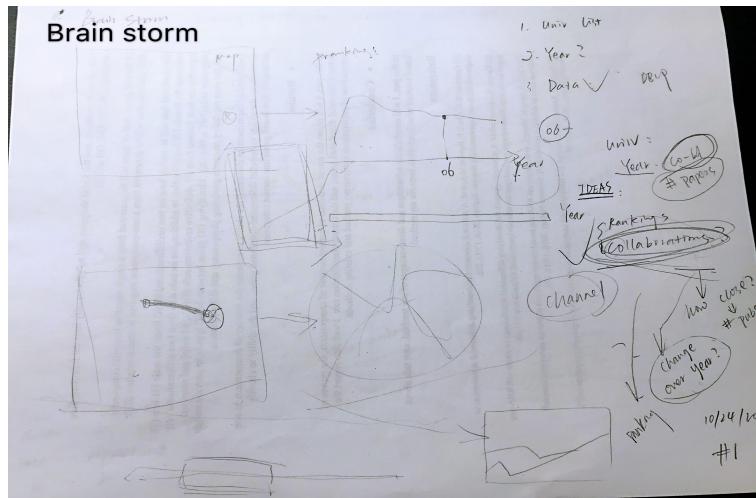
There are three main parts in our visualization:

1. The overview visualization of collaboration relationships for all universities, which will be realized by a world map, a connection graph, and some line charts.

2. Visualization of basic information for one chosen university, which will be realized by an information box, containing the name, the CS-ranking, the number of collaborators etc. of this university, also a pie chart and a line chart indicating the related information in specific fields (AI, Sys, AL, and Interdisciplinary areas).
3. Visualization of comparison information among chosen universities, which will be realized by a comparison table.

5.2 Sketches

We started our design with brainstorming ideas.

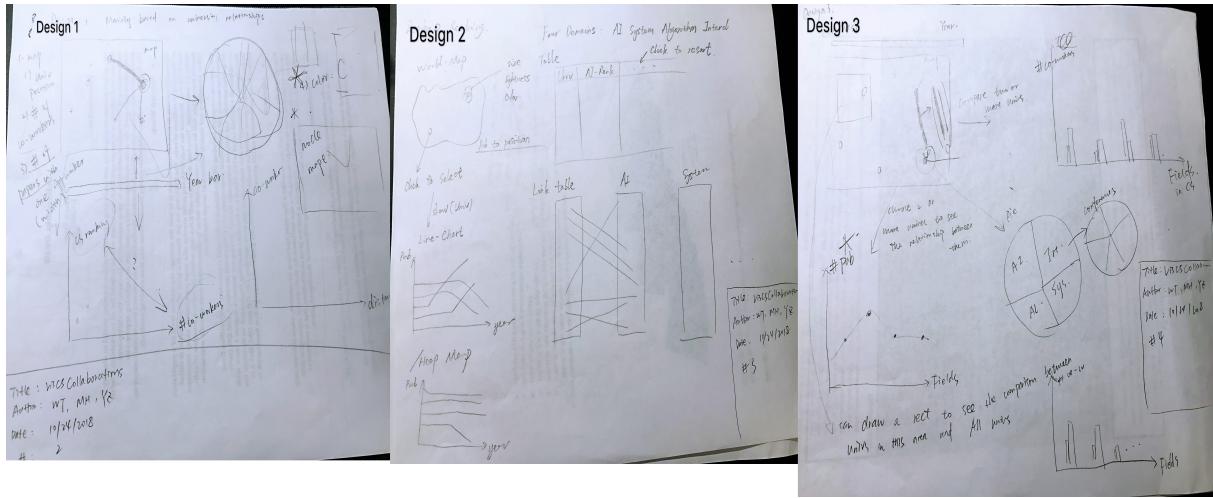


Then, we came up with three different design sketches.

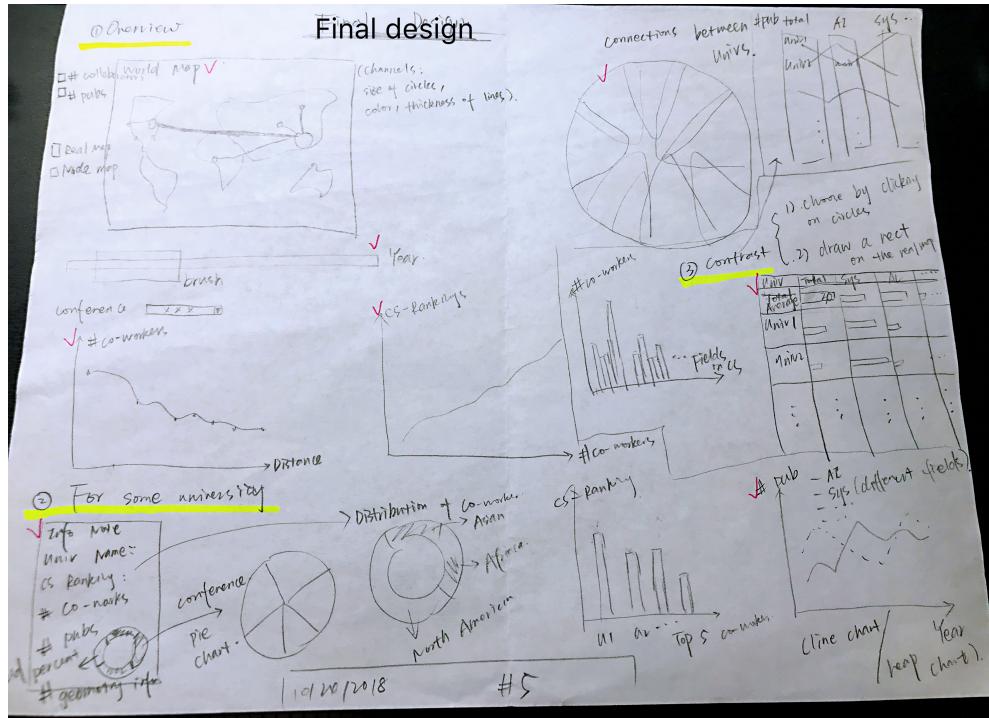
In our first design, we mainly focused on how to display the overview of collaborations between universities with the change of years. We came up with a real map with circles, and a connected graph to show the connections between universities. Also, we tried to show the relationship between the number of collaborators and other attributes by some line charts. We considered to add an interactive 3D scatter plot too.

In our second design, we mainly focused on displaying the collaboration relationships in some specific areas in computer science, including AI, systems, theory and interdisciplinary areas.

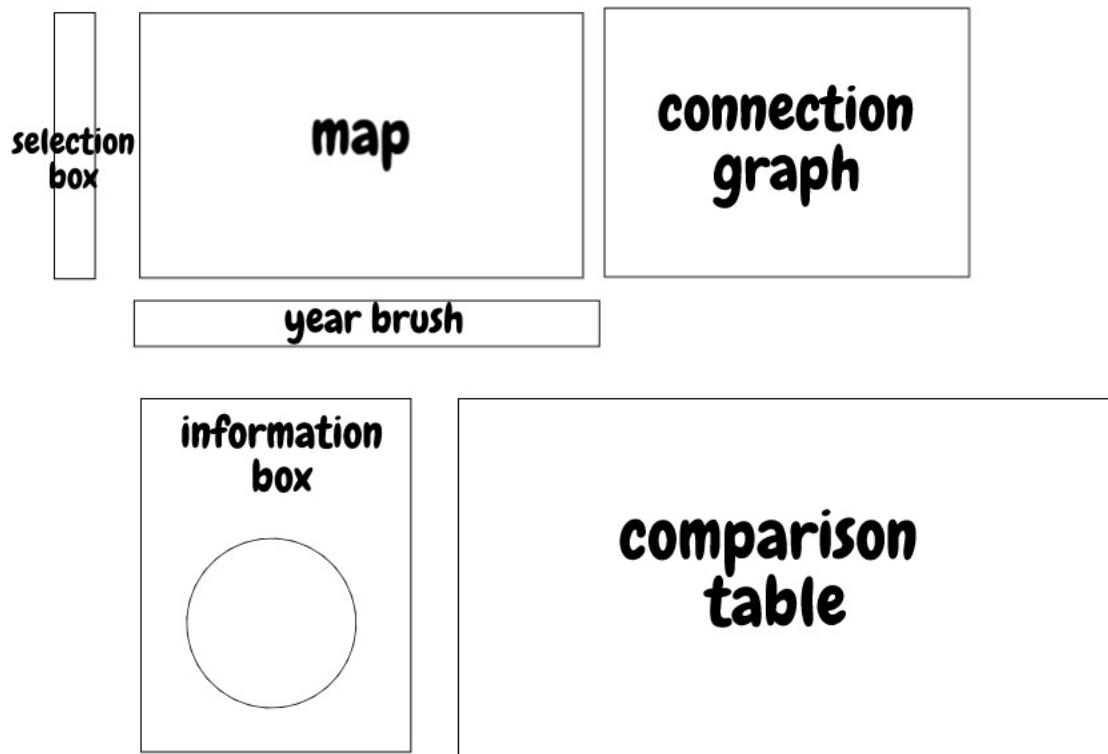
In our third design, we kept the world map, but mainly focused on representing the individual information and comparison information. Especially, we designed two ways of selecting circles. One is to click on the circle directly, and another is to draw a rectangle on the world map then circles in this rectangle will be chosen.



In the final design, we combined the advantages of these three designs and came up with a design including the three main parts stated above.



5.3 Prototype

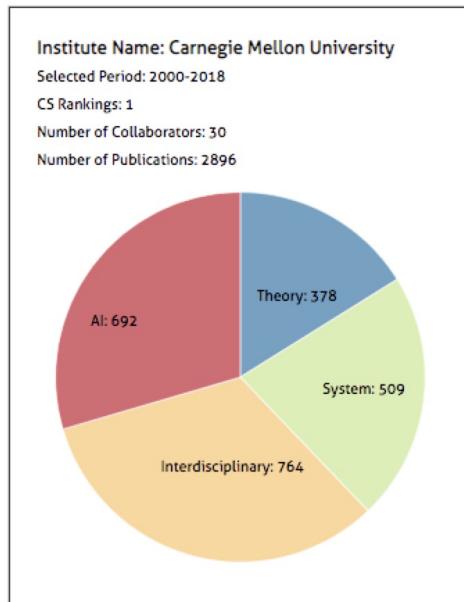


5.4 Implementation

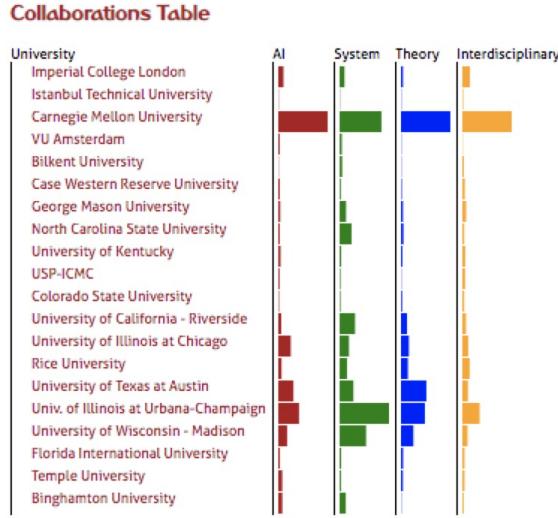
1. **The map:** After implementing the world map and placing institutions on the map as circles, we found that circles are so dense in the United States and European area that it is hard to select a specific circle. Therefore, we are considering to modify this view to display institutions only in the United States and their collaboration relationships with institutions inside and outside the United States.



2. **Information box:** This is an basic overview of a certain university, including university name, CS-Rankings, number of collaborators, number of publications. A pie chart to display the composition of its publications is also included in this box.



3. **Comparison table:** This is similar to the contrast table we built in hw5, which will show a bar chart of the number of collaborators in each cell. Clicking the headers of this table will sort the whole table.



6 Future Plans

Date	Goal	Task Assignment
11/9 - 11/15	Finish must-have features	1. Adjusting map (WT) 2. Information box (YZ) 3. Comparison table and Connection graph (MH) 4. Process book(YZ, WT)
11/16 - 11/22	Finish optional features	1. Parallel Chart (WT) 2. Line Charts (YZ) 3. Subfield Bar Chart (MH) 4. Process book (MH)
11/23 - 11/29	Test website and finish screen-cast	1. test website and debug (MH, YZ) 2. process book (WT) 3. screen-cast (WT, YZ)