Project Proposal

1 Basic Info

Project title: VisCsCollaborations

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

Project Repository: https://github.com/zhou325/dataviscourse-pr-VisCsCollaborations

2 Background 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.

At first, we focused on the global rankings in CS areas, but we felt like exploring the collaborations between universities is more interesting. Therefore, in this project, we are interested in visualizing the collaborations between universities using the DBLP dataset based on the different institutions, areas in the world, and other attributes. 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 our dataset, we define such relationship as the number of publications two institutions 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.

3 Project Objectives

VisCsCollaborations 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. VisCsCollaborations will show the rankings and relationships among the universities in the dimensions of location, topic and time, using groups of charts to. According to their data scales, the charts can be divided into three groups: world view, contrast view and certain university view.

World view:

VisCsCollaborations will posit every university on a world map so that it will be convenient to watch and discover the geographical distribution of research affiliations in the field of Computer Science. Specifically, VisCsCollaborations will take the user' selection as the period of interest and present the corresponding data. By this method, user will be able to see the development of the universities research outputs. And it will be easy to find out how a certain research affiliation was doing among its

peers. Meanwhile, we can quickly tell which universities are the top ones that are serving the CS research in a certain period.
On the other hand, VisCsCollaborations will link universities that involve cooperation in CS publication. With this tool, the users can easily recognize how many cooperative companies the universities own and among these who is the closest one. Also, VisCsCollaborations could help discover the correlations between the overall CS rankings, closeness geometrically and cooperation variety.

Contrast view:

VisCsCollaborations 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. VisCsCollaborations use 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.

Certain university view:

If the user is interested in learning more information of a certain university, VisCsCollaborations allow 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

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.

We can download the DBLP data from https://dblp.uni-trier.de/xml/ and there are two files need in parsing data which are dblp.xml.gz and dblp.dtd files. Different from first dataset, second data is already collected and processed by one of our group member due to a person interest.

5 Data Processing

Due to size of dataset, we would like to pick only a subset of this DBLP dataset in our visualization project. Before picking up an appropriate subset which we may take a further consideration later, such as top 100 universities in certain research areas based on the number of publications in Journal Article, Book, and Conference Papers, etc., we need to parse the xml file into a json/csv/txt file firstly. We will write a DBLP parser to extra useful features based on our need to convert the dataset into json file. We then will decide which part of json data we may use in the data visualization.

6 Visualization Design

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.

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.

7 Must-Have Features

1) **A real-world map:** Universities are represented by circles. The position of circles on map indicating the location of universities in the real world. The size of circles indicates the number of publications or the number of total collaborators in a specific period (this can be realized by a checklist combining a year bar). When clicking on a circle, connections between this university and other universities will be displayed.
2) **A connection graph:** Universities are placed on a large circle, and the relationship between universities is represented by a line connecting two universities.
3) **Distance vs Number of Collaborators:** indicating the average number of collaborators in different distances.
4) **Number of Collaborators vs CS-Rankings:** indicating the relationship between the number of collaborators and CS-rankings
5) **University Basic Information**
    i) **Information Box:** including university name, CS-Rankings, number of collaborators, number of publications and the location of this university.
    ii) **Donut Chart:** indicating the proportion of collaborator amount in some specific areas.

iii) **Year vs Number of Publications:** indicating the total number of publications in each year in total and in different areas.

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

8 Optional Features

1) **Sorted Bar Chart:** A bar chart displaying the publication amount of the best universities from top to down according to the ranking.
2) **Contrast Graph**
   i) **Line Chart:** presenting and contrasting several universities' yearly publication.
   ii) **Subfield Bar Chart:** displaying the subfield publication of selected universities with bars.
   iii) **Parallel Chart:** showing the university rankings in total and subfields. There will be lines to link the same university in different ranks.

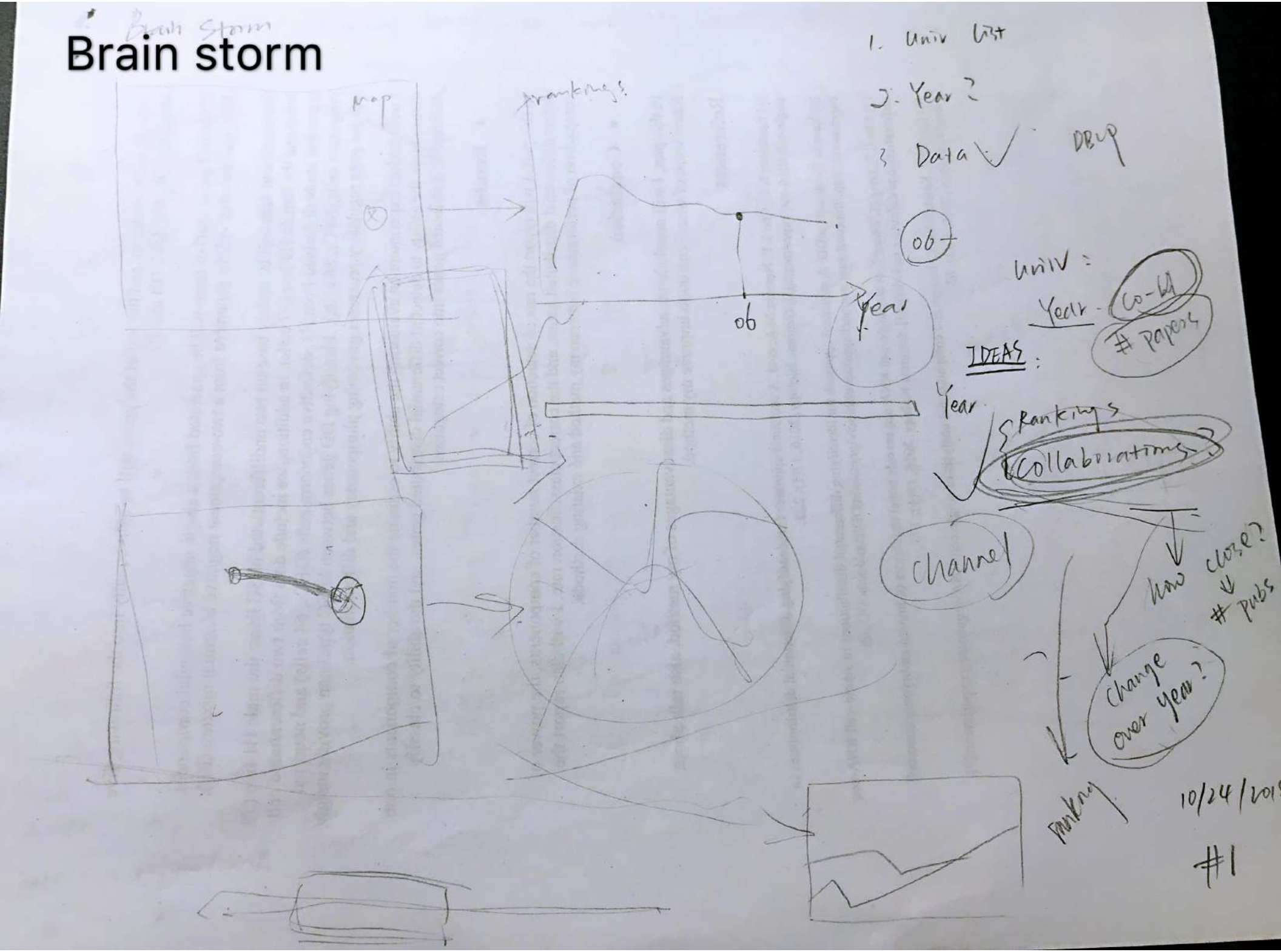
9 Project Schedule

We will meet every Wednesday to review the work of last week.
The task assignment will be updated every week after group meeting.

(WT: Wenzheng Tao, MH: Mingxuan Han, YZ: Youjia Zhou)

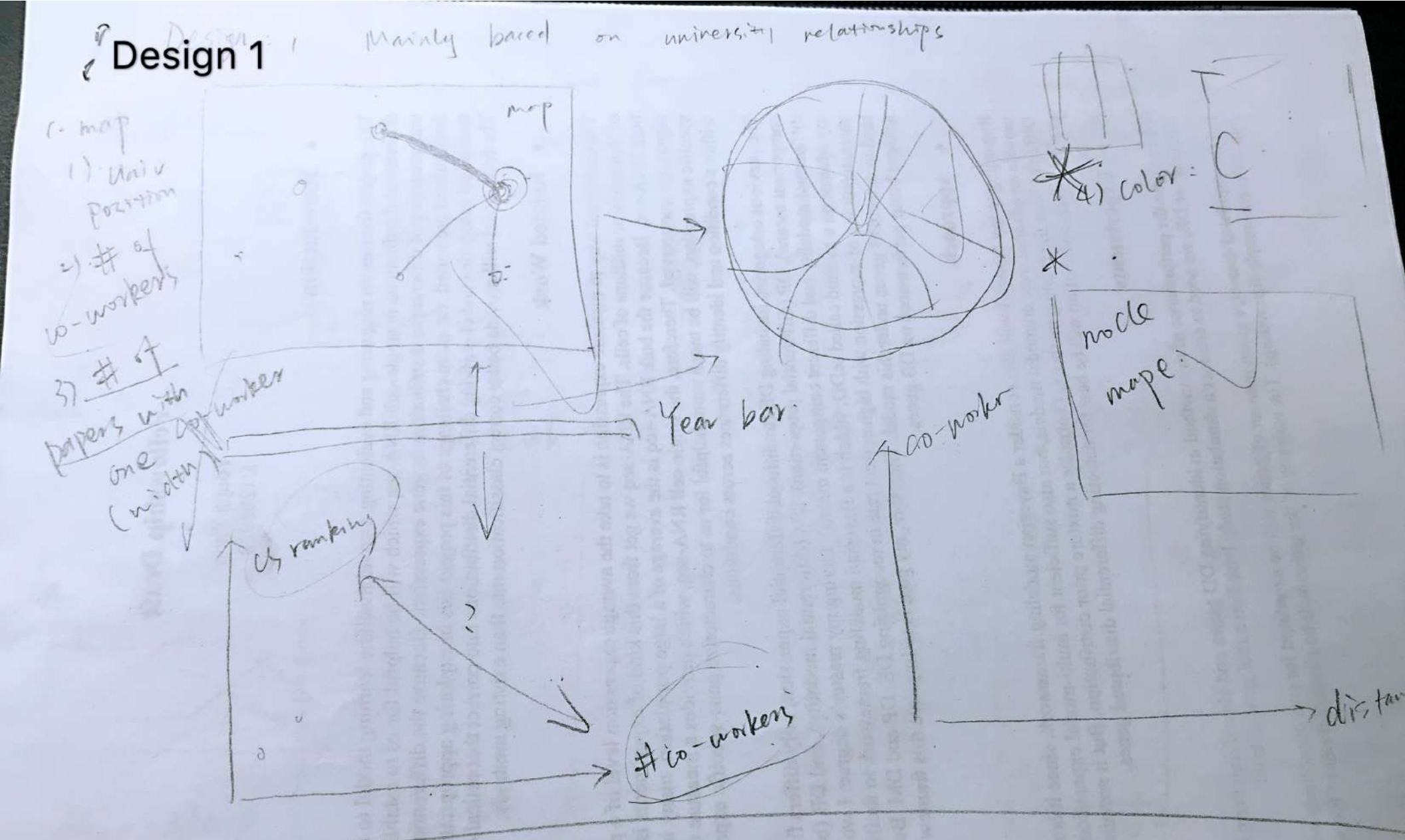
Date	**Goal**	**Task Assignment**
Week 1 (10/26 - 11/1)	Initialize visualization	1. data cleaning (WT) 2. layout design (MH) 3. data structure design (MH) 4. process book (YZ)
Week 2 (11/2 - 11/8)	Build a website demo using sample dataset	1. implement each part with the sample dataset (WT, MH, YZ) 2. process book (MH)
Week 3 (11/9 – 11/15)	Finish must-have features	1. implement must-have features with the whole dataset (WT, MH, YZ) 2. process book (WT)
Week 4 (11/16 – 11/22)	Finish optional features	1. implement optional features with the whole dataset (WT, MH, YZ) 2. process book (YZ)
Week 5 (11/23 – 11/29)	Test website and finish screen-cast	1. test website and debug (WT, MH, YZ) 2. process book (MH) 3. screen-cast (WT, YZ)

Brain storm


Brain Storm
1. Univ list
2. Year ?
3 Data ✓ DBLP
map
rankings
06
Year
Univ :
Year - Co-Ld
papers
IDEAS :
Year
Rankings
Collaborations
channel
how close?
pubs
change over year?
ranking
10/24/201
#1

Design 1

Mainly based on university relationships



Title : UCSC Collaborations
Author : WT, MH, YZ
Date : 10/24/2018
: 2

Design 2

Four Domains: AI System Algorithm Interd

Click to resort.

World-Map

size
lightness
color

Pub to position

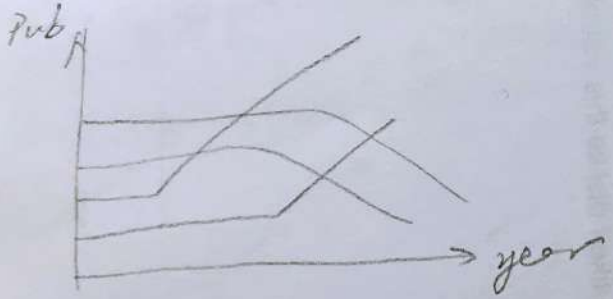
Table

Univ	AI-Rank	. . .

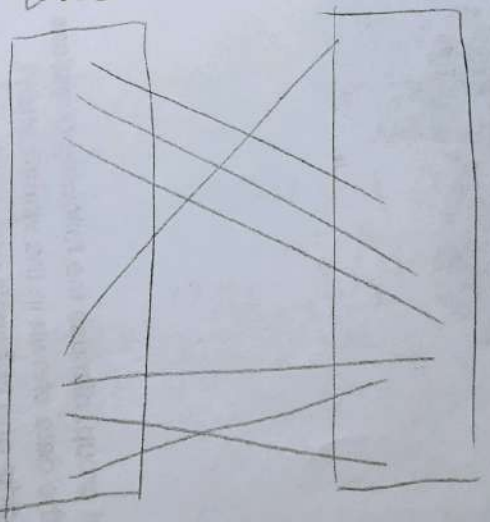
Click to Select

show (Univ)

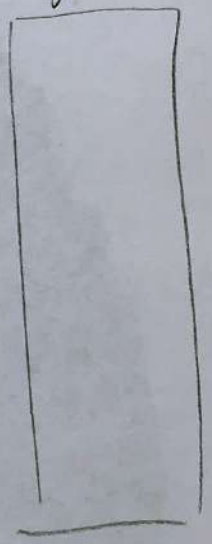
Line-Chart



Link table

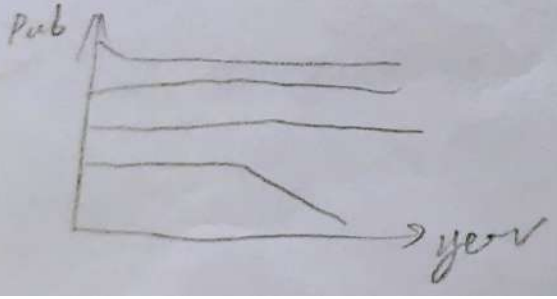


System



. . .

Heap Map



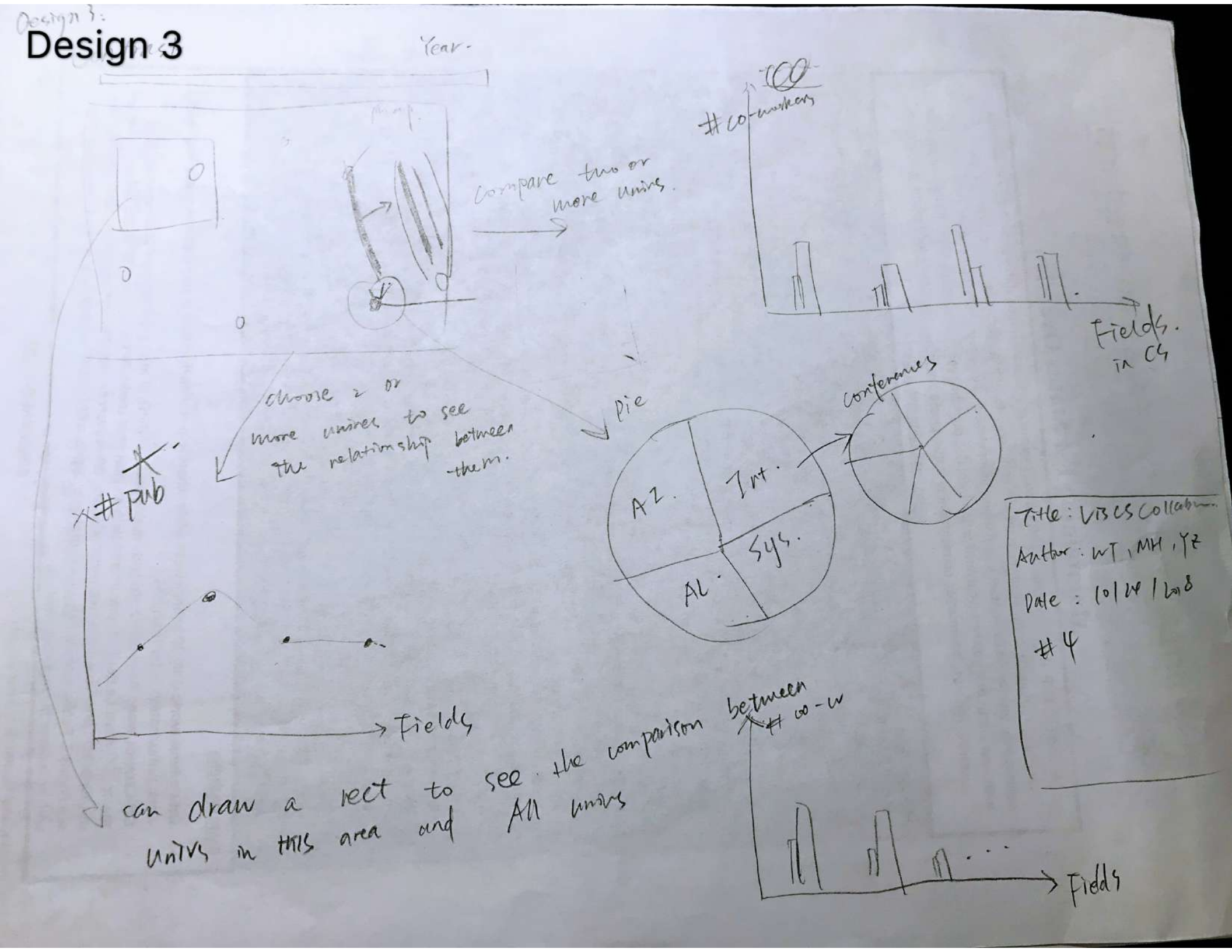
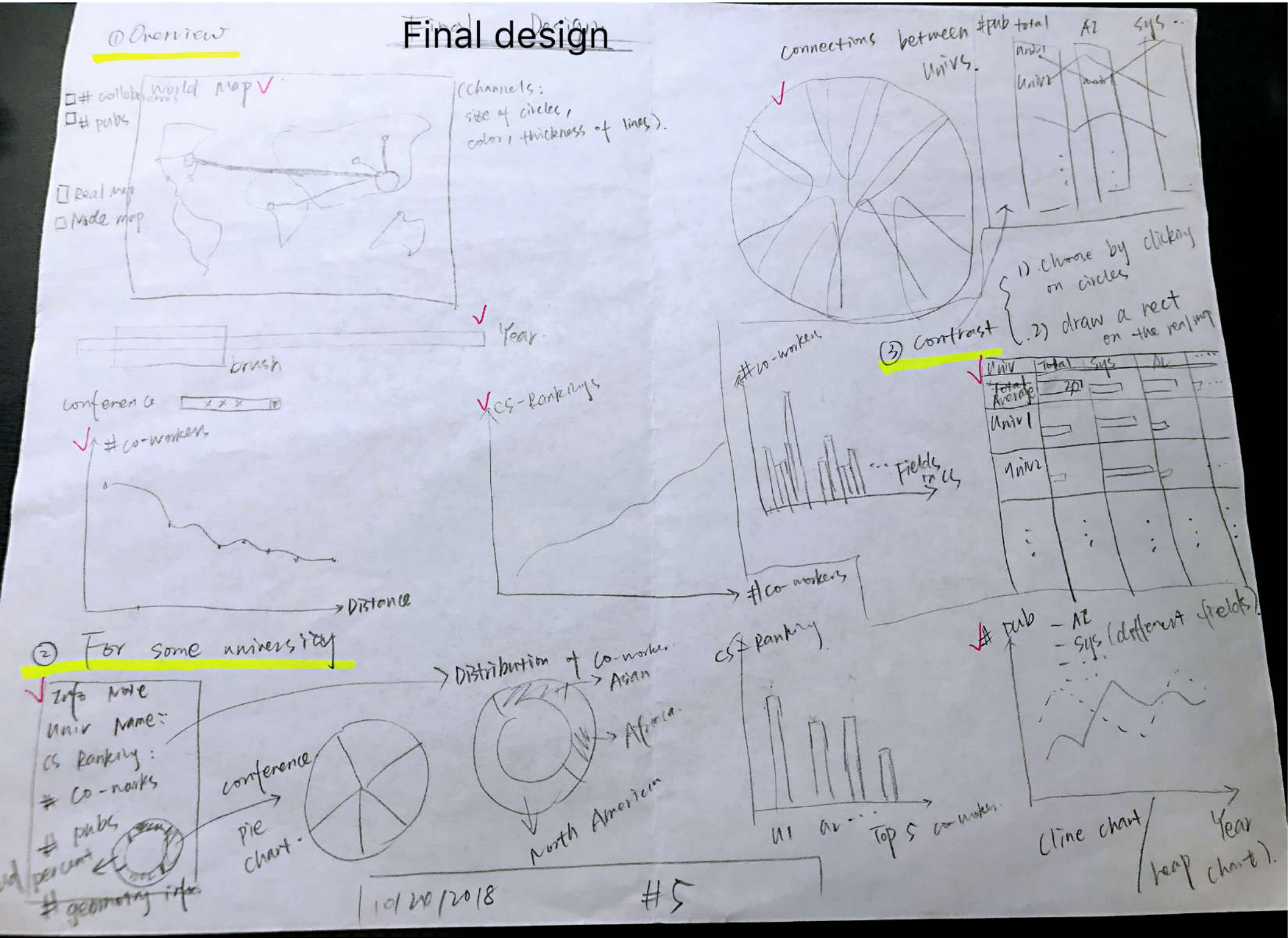
Title: VISCSCollaboration

Author: WT, MH, YZ

Date: 10/24/2018

#3

Design 3


Year.
compare two or more univs.
co-workers
Fields. in CS
choose 2 or more univs to see the relationship between them.
pie
conferences
AI.
Int.
Sys.
AL.
pub
Fields
Title: VISCS Collab...
Author: WT, MH, YZ
Date: 10/24/2018
#4
can draw a rect to see the comparison between univs in this area and All univs
co-w
Fields

Final design