CS 6630 Project Proposal: Company Links

Alec Nelson	Yu Zhu	Yuanyuan Tong
u0741446	u1211862	u1142943
u0741446@utah.edu	yu.zhu@utah.edu	yuanyuan.tong@utah.edu

1 Background and Motivation

We were inspired to pursue research in this area based on our team members' interests in financial markets and US businesses. Now more than ever, trading on stock exchanges happens in an automated manner. Many analysts build complex models to identify trading opportunities built from data-based approaches. Predictions of a company's future performance can be derived based on a variety of factors. There exists a wealth of data on public US companies, and therefore good potential to unearth surprising or unexpected relationships that exist within the landscape of large businesses and firms.

2 Questions and Objectives

With the capability of our data, we are trying to answering the following questions.

1. How is the geospatial distribution of public companies in the US like?
2. What kinds of connections exist between publicly traded companies, and can these connections be correlated to stock prices?
3. If a board member leaves their job, do they tend to stick around and accept a new job in the same region? Do they prefer to stay in the same industry?
4. Is the past education (or lack thereof) of big company board members indicative or predictive of their respective industrial sectors?
5. Does a company (or an industry) have an inclination to recruit board members with certain education background?
6. What kinds (if any) of cross-industry connections exist between large US companies?

More questions could be answered with the deepening of our research as well. By answering these questions we hope to uncover non-trivial relationships and links between large US companies. We want to enable interactive exploration of the landscape of big players in the US economy, and hope these derivations will be useful in developing financial models and analyzing US companies.

3 Data

Our raw data are from two commercial databases subscribed by the Eccles' Business School. The first database, called "BoardEx", is a dedicated database covering compensation, employment, and relationship data for 20,000+ companies globally from 1999-present. We obtain profiles, education background and employment history for all the individuals that have served in any US listed company. We also obtain links between companies established by these individuals. For example, if a person is a board member in company A before he is hired by company B, then we say a link has been established by this person.

For company related variables such as market capital, industry, SEC category code, we refer to the "CRSP (Center for Research of Stock Price)" database, one of the most authoritative in academic research.

4 Data Processing

The data processing goes as follows. First, we download the raw data, which may contain many observations and variables that we may not use. In the second stage, we group these variables into five tables according to their inherent relationships and remove unnecessary observations/variables/outliers. The descriptions of the five resulting tables and the specific variables inside each table are shown in Table 1.

Table Name	**Description**	**Table Fields**
ind_prof	Individual Profiles	name, name_id, date_of_birth, date_of_death, age, gender, nationality
ind_edu	Individual Education	name, name_id, education_type, university, university_id, qualification (e.g., BA, PhD), qualification_awarded_date
ind_emp	Individual Employment	name, name_id, company, company_id, role, role_start_date, role_end_date, if_in_board, if_in_supervisory
firm_prof	Firm Profiles	company, company_id, ticker, market_cap, industry, industry_code
firm_net	Firm Links (each row records a link from "from_company" to "to_company")	from_company, to_company, name, role_in_from_company, role_in_to_company, role_in_from_company_start(end)_date, role_in_to_company_start(end)_date

Table 1: Final data formats after data processing

5 Visualization Design

The overall design of our visualization will be split into two major components, namely the **map** and the **secondary view**. The map will be the focus of our project and will be positioned on the left-hand side of the page. Since the map is best suited to display geo-spatial relationships, we will include more detailed aspects of the data in the secondary view. The secondary view will be placed to the right of the map and will outline relevant features of our data. We will allow the user to select a company on the map, and information about this company will appear in the secondary view. Sketches are shown at the ending of this document.

6 Must-Have Features

Perhaps the most essential component of our visualization is the US map. The map will act as a base-layer that additional features will be built on top of. This map must be zoomable and scrollable, and should properly display state borders. This map must be geographically accurate, as we will be adding geotagged information on top of it.

The next essential aspect of our visualization (arguably equally important as the map) are the company tags themselves. At this point we are not entirely sure how we will be displaying them on the map, but right now are leaning towards circles whose sizes are based on that company's market cap.

Next, we will implement the core functionality of our visualization: showing the important cross-company relationships from specific dimensions. The personnel migrations between companies would presumably be the strongest links we could find with our data. The relationships would be represented by a adjacency matrix, a chord diagram or just normal networks. We would preferably display such relationships on top of the map view by incorporating the geographical information.

We may also dive into a company's profile and make inner-business explorations. This might be a pie chart showing board member constitution, e.g., in terms of education background. And this view will only appear in response to the user's demand.

7 Optional Features

If we could find some salient and interesting patterns about the relationships between the companies, we might do storytelling by first navigating the audience from certain viewpoints. The audience could build up their knowledge of the data and the usage of this visualization thus being able to explore the visualization themselves.

We would like to have multiple views of our visualization, besides just the map. For example, it would be great to be able to brush over particular regions on the map and expand them in more detail in a separate window. Another view could display statistical data, perhaps through the use of bar charts or line plots. We have also discussed the possibility of being able to search our data through an input field, however will only explore this if time permits and the essential features are completed.

We have also contemplated the idea of introducing time to our visualization. This introduces another level of complexity, and a lot more data collecting that would be necessary. However, doing this could allow for some more insightful interpretations of our data, as well as more storytelling capabilities. The time variable remains auxiliary at this point - we will have a better idea if we want to incorporate it into our visualization as we begin exploring our data within our first prototype.

Other subtle features, like showing the company profile in tooltips when hovering over it on the map, are also in our consideration if time permits.

8 Project Schedule

Week 1 (10/26 - 11/1)

Work together to narrow down project scope. Populate GitHub repository with core project components (Yuanyuan). Gather and clean all data necessary for project(Yu). Begin exploring data on interactive US map(Alec). Write in process book.

Week 2 (11/2 - 11/8)

Complete visual prototype (interactive map that displays companies by location and market cap, as well as visual links between company staff). Ensure prototype meets milestone requirements. Write in process book. Members will work together to make the prototype run.

Week 3 (11/9 - 11/15)

Visualization core components are working. All basic functionality exists. Begin implementing secondary features such as alternate views, brushes, and additional information boxes. Each member would be free to explore 1-2 alternatives, and the most interesting one will be integrated into the final visualization.

Week 4 (11/16 - 11/22)

Have visualization completely finished. All moving parts are working correctly. Incorporate storytelling. Changes beyond this point are stylistic in nature.
Yuanyuan may focus on the story telling part. Alec and Yu will work together to debug the overall website.

Final Days (11/23 - 11/27)

Finalize CSS styling. Improve efficiency where possible. Make sure HTML is valid. Ensure code commenting is sufficient and clear. Test code extensively on various machines and in different browsers. Write in process book. Complete screen-casts. Complete peer assessments. Ensure project requirements are fulfilled. Organize all material and submit assignment.
At this stage, most of the work will be the refinements and we will mainly work as a group.

Sketches

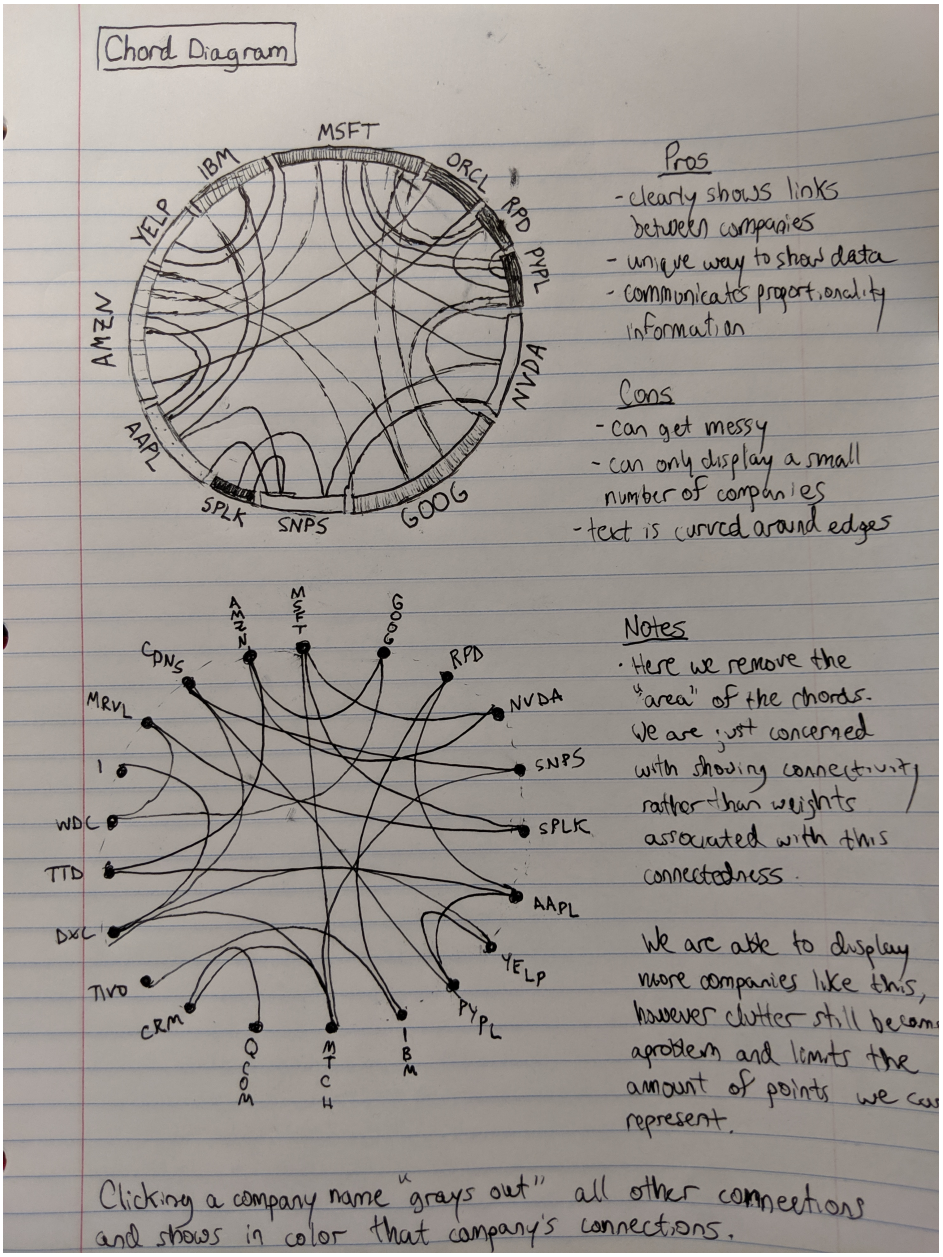

Figure 1: Initial Design 1: Chord Diagram

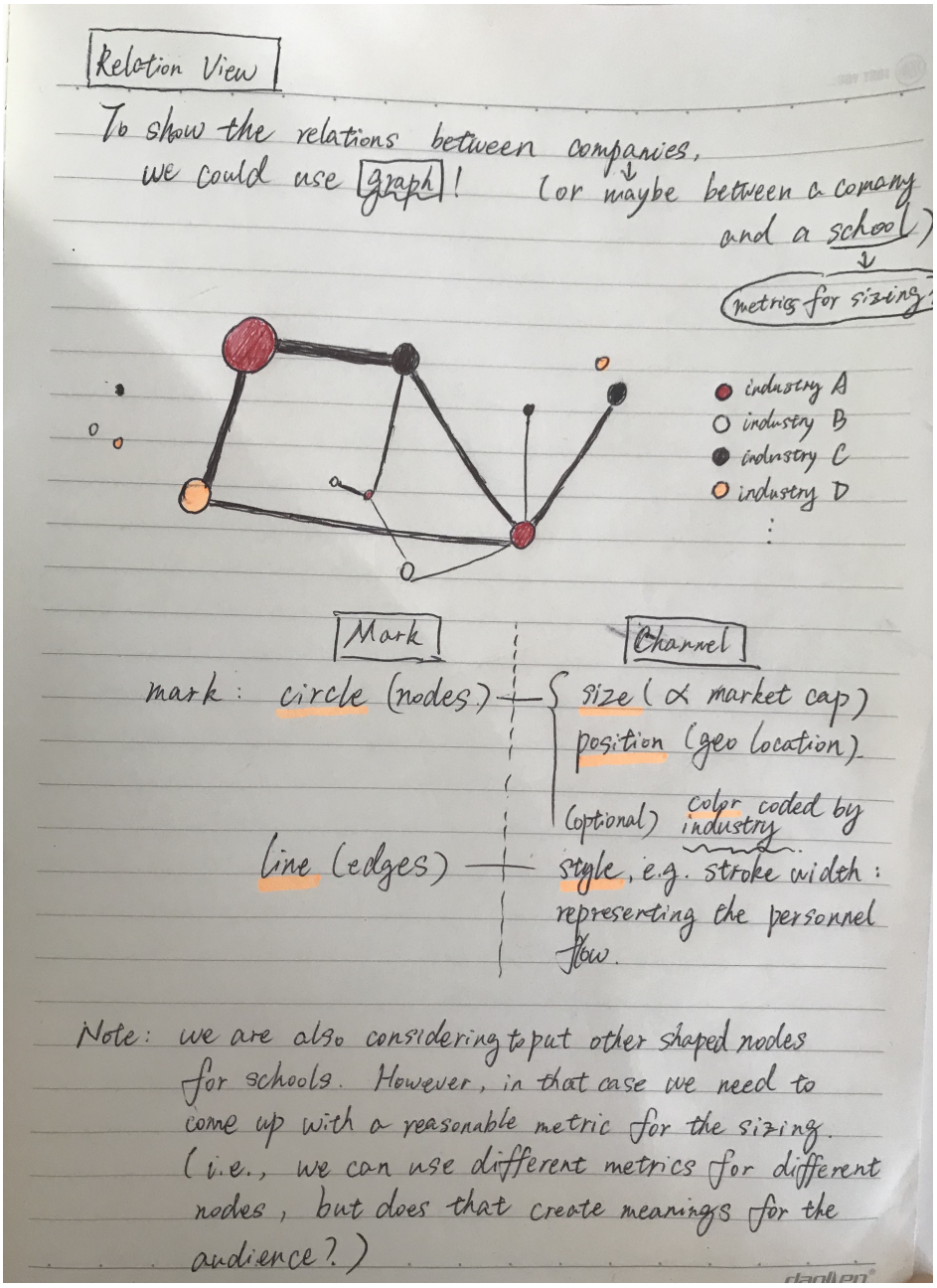

Figure 2: Initial Design 2: Network Graph

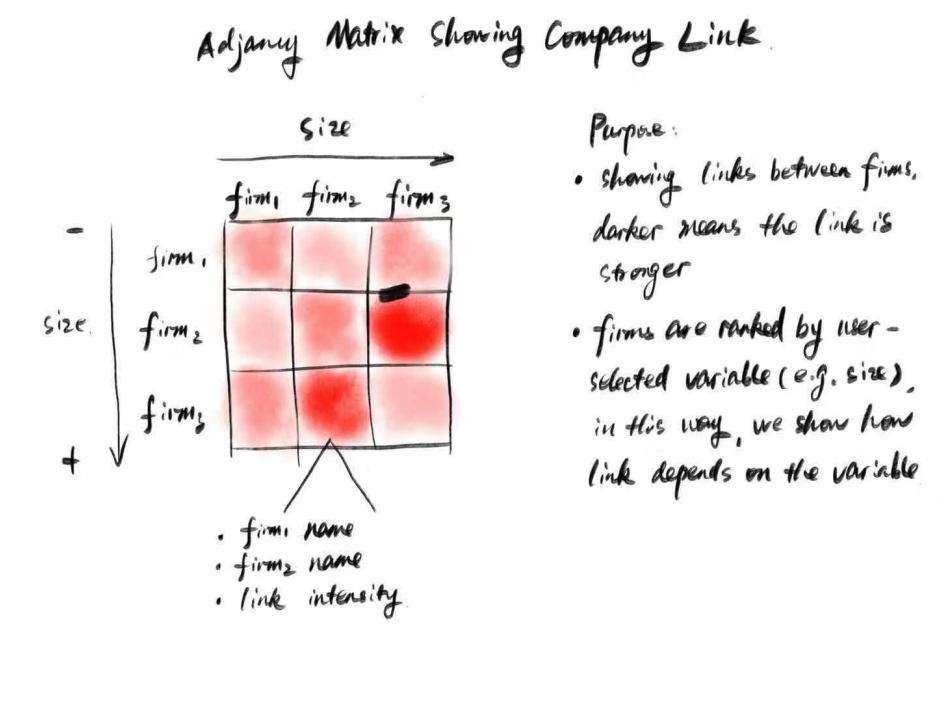

Figure 3: Initial Design 3: Adjacency Matrix

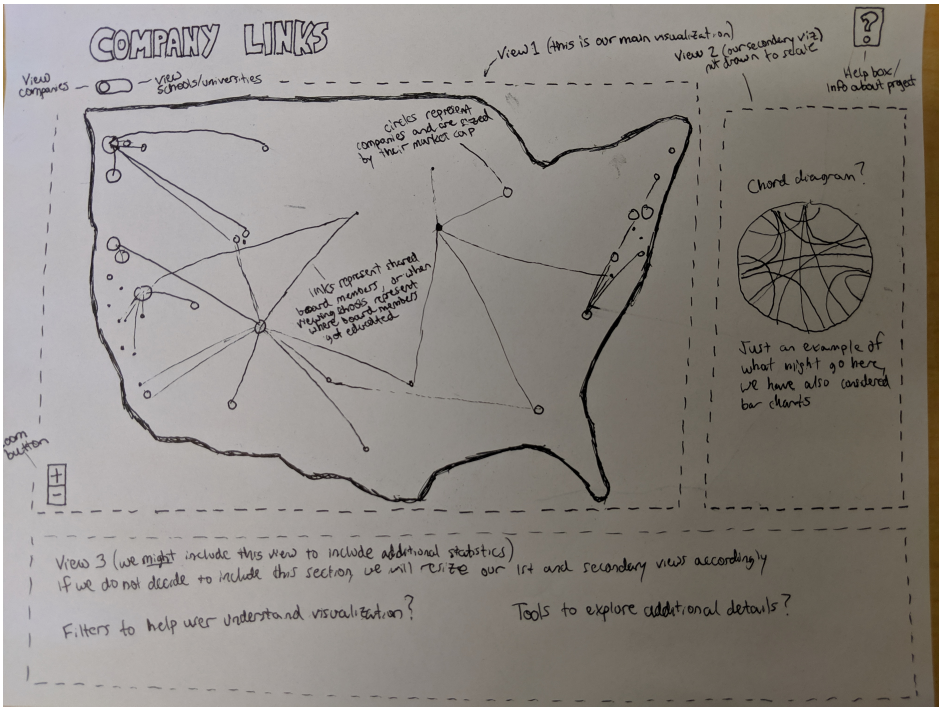

Figure 4: Final Design: Main Map View

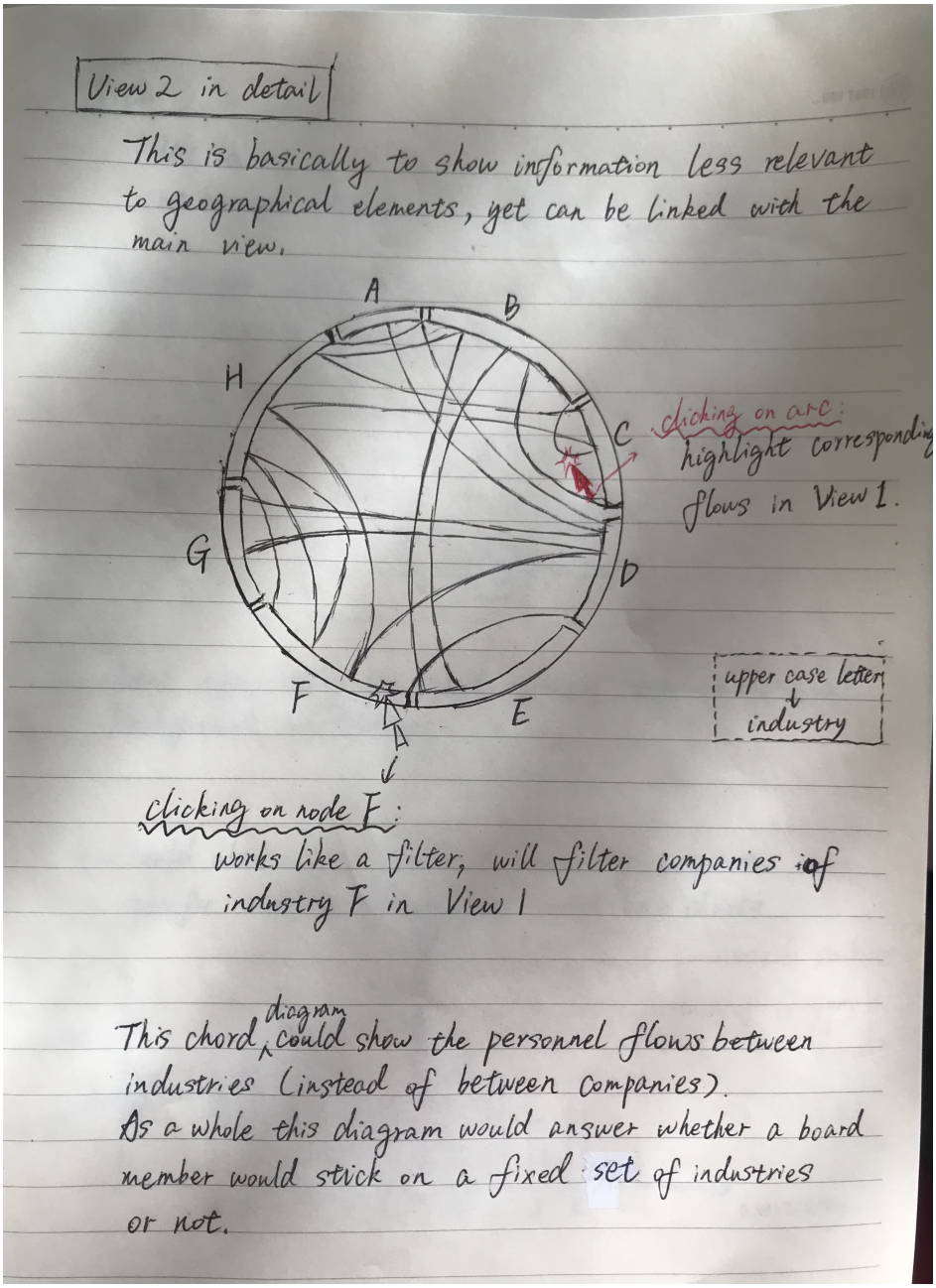

Figure 5: Final Design: Secondary View

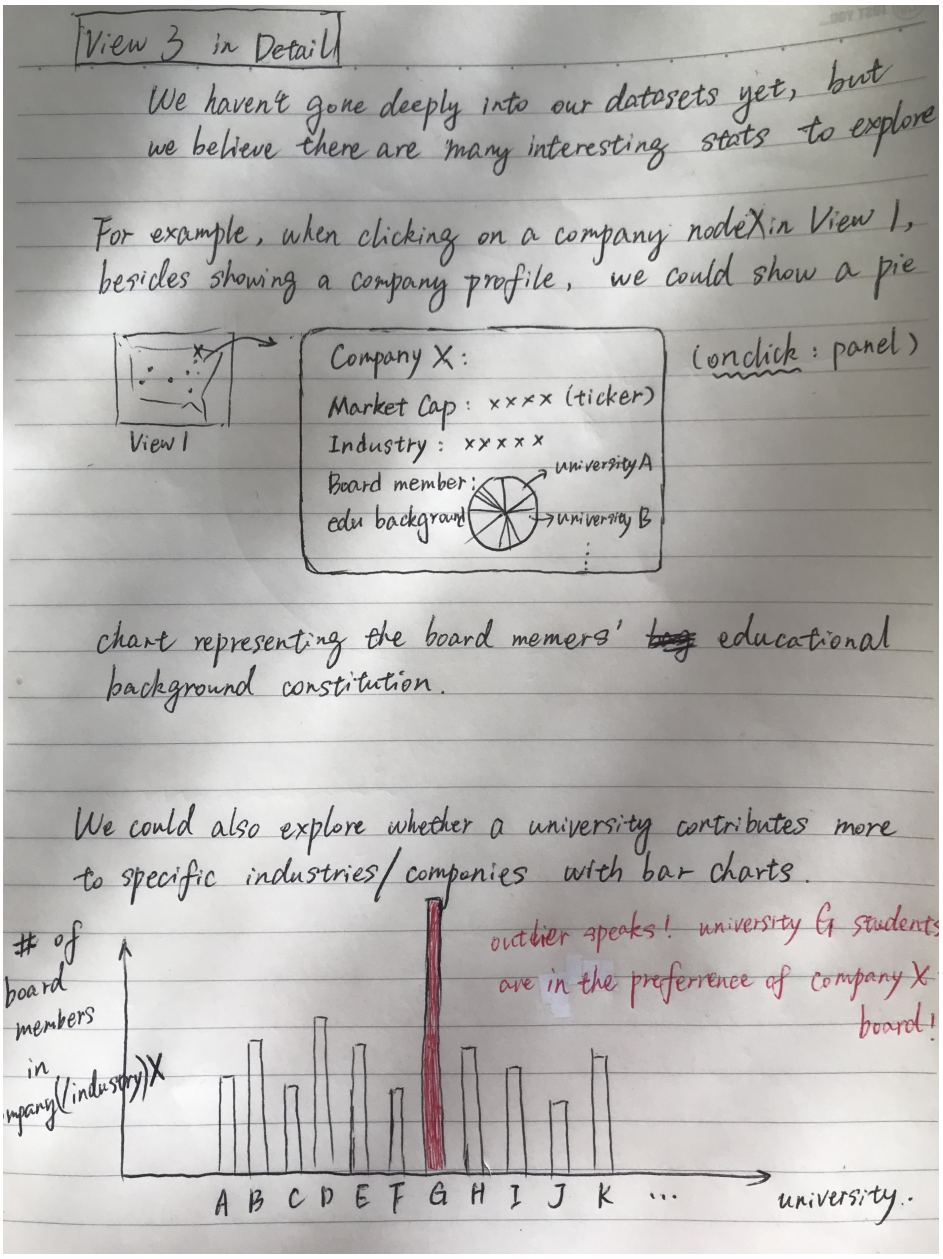

Figure 6: Final Design: View 3