Academic Visualization
An Insight Into Academia

Changfeng Liu

June 20, 2016

Abstract

In this report, I will introduce my project *SchoolDrag* of EE327, Wireless Communications and Mobile Networks. *SchoolDrag* is a academic visualization prototype to demonstrate the publication numbers and developing trends of different academic organizations. This report consists of four sections, namely background, work review, SchoolDrag implementation and other visualization work.

1 Background

1.1 What is visualization?

The very first question is what is visualization? According to Wikipedia[1], information visualization is the study of interactive visual representations of abstract data to reinforce human cognition. In my opinion, visualization is an easy way to understand the world.

1.2 Why academic visualization?

The following question is why we need academic visualization. Professor Wang said: "We have got massive academic data, which is like raw food materials. What we should do is to cook delicious dishes!" How to cook delicious food and be a good chef? Of course we won't go to Xindongfang, AceMap is the right answer!

In AceMap, we have the chance to see how a delicious dish is made. Firstly, crawling and data cleaning are needed to obtain available data. Then database is used to store the data. After that, some data mining techniques are used to cook the data. Finally, information visualization will present the data to the consumer. A picture is worth thousands of words, a good visualization will help the user to understand the complex structures between papers and reveal the hidden trends in academia.

Since visualization is important, can we use visualization from other websites? The answer is frustrating: because traditional academic search engines provide few visualizations. For example, when I type in data mining, Google Scholar can only return a list of papers without any visual information. This can be a disaster for a recruit researcher.

2 Work Review

On winter vacation, our paper [2] was accepted by WWW BigScholar workshop and we submitted two patents related to academic visualization.

We mainly focus on local network analysis and hierarchical clustering among different papers in that work. Figure 1 shows some visualization pictures from paper [2].
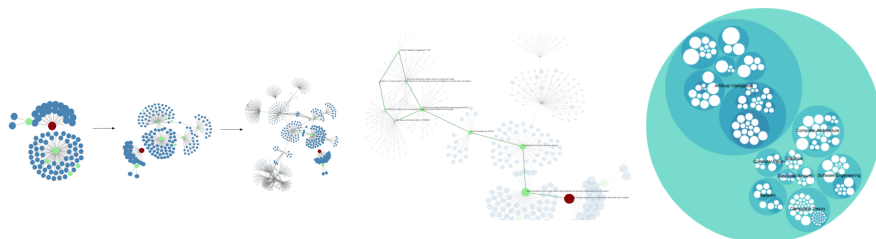

Figure 1: Some visualization prototypes of paper [2]

As I have mentioned, the visualization of our previous work focused on demonstrating the pictures of papers. However, beyond paper plus paper, we can see more through visualization. For example, paper plus university, paper plus timeline and paper plus location. The combination of different types of entities will make some really informative and interesting visualization.

3 SchoolDrag Implementation

This semester, our team implemented three visualizations. They are about the academic organizations, geographical locations and hierarchical structure. This report only focuses on SchoolDrag.

3.1 Motivation

The motivation of SchoolDrag is following: How to select good graduate schools in a certain field as most of us will face this problem several months later, how to find out the developing trends of different universities and which university is the groundbreaker and which is the follower. Unfortunately, few academic search engines can answer these questions.

3.2 Visualizaiton tool selection

So we have to do this work by ourselves. The first thing we care about before conducting visualization is which tools to choose. Since we want to design new styles and make them available on the Internet, we consider four aspects. Open-source, easy to code, community support and web-based. As a result, we choose *vega* and *D3* as our tool.

3.3 Handle the data

What we want to get from the database and show on the Internet are three things:

1. To know the numbers of published papers of every organization in every year.

2. To get the values of corresponding authors of every organization in every year.

3. To demonstrate the developing trends and academic potential of different organizations in an interactive way.

Of course, *vega* and *D3* are not enough for our visualization. We use *MySQL* to manage and retrieve data, *C++* and *Python* to do the computation, *JSON* to store the data and *batch files* to do the repeating tasks. The relation between these tools is shown in figure 2.
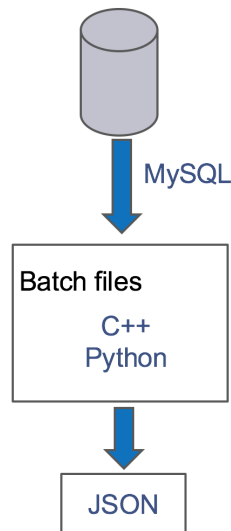

Figure 2: Relation between tools.

We have got two fields information according to CCF recommendation list, namely Network and Artificial intelligence. Recently, we have calculated ten L1 or L2 fields according to MAG dataset, for example, computer security, computer vision, embedded system and so on.

3.4 Demo

SchoolDrag's interface is shown in figure 3. As you can see, there are several useful features of our prototype:

1. Each circle stands for an academic organization.

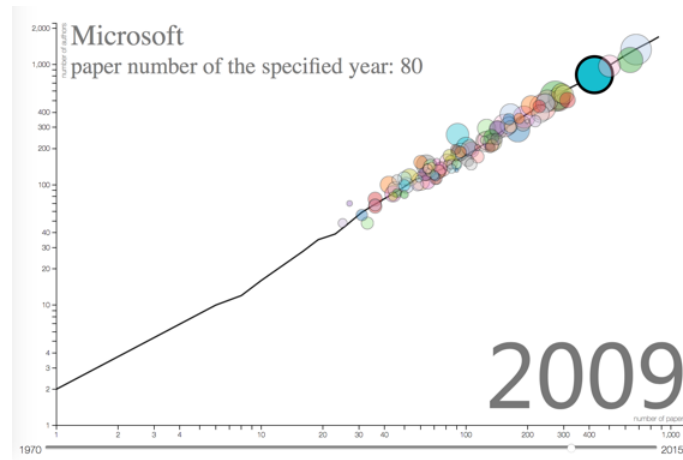2. The more published papers an organization has in the specified year, the bigger the circle is.

Figure 3: The interface of SchoolDrag

3. The curve shows the developing trends of the total number of published papers and authors of an organization.

4 Other Visualization Work

I also constructed a visualization for our paper for *ICDM 2016*. This demo presents the relationship between academic papers.

References

[1] Wikipedia. *Information visualization.* https://en.wikipedia.org/wiki/Information_visualization

[2] Tan, Zhaowei, Changfeng Liu, et al. *AceMap: A Novel Approach towards Displaying Relationship among Academic Literatures.* Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, 2016.