Wireless Communications and Mobile Internet Project Report

Tang Weijie
School of Electronic, Info. & Electrical Engineering
Shanghai Jiao Tong University
Shanghai, China 200240
Email: tang_wei_jie@126.com

***Abstract*—This report mainly introduce the work I have done in 2016 spring semester for the course project—AceMap: Academic Information System. Specifically, it includes the whole process of realizing the feature on AceMap of 22000+ paper maps and 2200+ author maps, from extracting data out of Microsoft "mag" database through generating graphs with Gephi visualization software to writing the webpages for maps. All these maps are already online, making it convenient for users to view academic data and greatly improving the map feature on AceMap website.**

I. Introduction

AceMap is a novel academic information system put forward by professor Xinbing Wang. Before this semester, the website has already online with its basic search function. Users can type in any academic keywords of their interest and get related search results, including papers, authors, affiliations, fields of study and other academic information.

In this semester, my teammates and I focus on the primary feature of AceMap—the academic maps. In general, the maps currently on AceMap fall into two categories: global and detailed. The global maps show the overview relationship between all the academic information in the database. My teammate has generated two global maps: topic map and affiliation map. The former shows basic information of main topics in clusters; the latter displays academic influence of different affiliations on the global map frame. The detailed maps show more academic information within a relatively small range. I have generated two kinds of detailed maps: paper maps and author maps. More than 22000 topics have their unique paper maps; each shows the basic information of topic and the paper network. More than 2200 affiliations have their unique author maps; each shows the basic information of affiliation and the author network.

Furthermore, all maps online have several user interaction functions. While zooming in and out, labels of the nodes on maps will be displayed or hidden. By clicking each node, users will go to corresponding webpages. By hovering each node, detailed information will be shown in a floating window on the right side. There are still other functions waiting for you to explore. Certainly, all maps are linked together. Feel free to jump between maps for better experience.

The following part of this report is organized in the processing order of generating maps. In section II, it illustrates how to extract needed information. The details about generating and analyzing maps are explained in section III. After demonstrating user interaction in section IV, there is the section V about future work. The report is finished with a conclusion and an acknowledgement.

II. Play with Database

All the maps are generated from entries of data in database. So the very first step is to play with database. Though in this semester, I generated two kinds of maps, the data preparation process of them are similar as shown in Fig. 1.

Paper Maps	Author Maps
Connect database in Java	
Choose fields	Choose affiliations
Extract papers	Extract authors
Extract references	Extract coauthors
Filter	
Generate .gml	
22000+	2200+

Fig. 1. Data preparation steps.

Considering the large number of maps, I use Java rather than other GUI interfaces to do the process. After connecting database in Java, three sql queries are executed to obtain the data for paper maps and author maps respectively. The first step is to decide in which research fields or affiliations I am going to generate maps. The second step is to extract papers or authors information as the graph nodes. The third step is to extract references or coauthors information as the graph edges. After filtering the raw data, write all the data into a new gml graph type file.

III. Play with Graphs

The purpose of this process is to use the gml files to generate those maps on the website, which is in svg file type.

A. Generating Maps

There are several steps to generating svg maps from gml files, as shown in Fig. 2. After importing gml file into gephi workspace (also in Java), it will give a nearly random black graph. First it should be filtered that nodes with zero degree

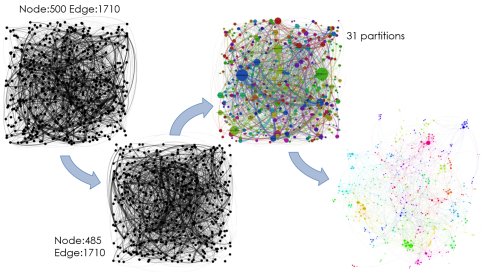

Fig. 2. Maps generating steps.

are not welcomed. Then implement relevant algorithms to set attributes of the graphs according to the network information, and get colorful graphs. Finally, run proper layout algorithms to layout graphs and get the good looking maps.

B. Paper Maps

Each topic fields have a paper map. Here I choose the research field of machine learning as an example shown in Fig. 3.

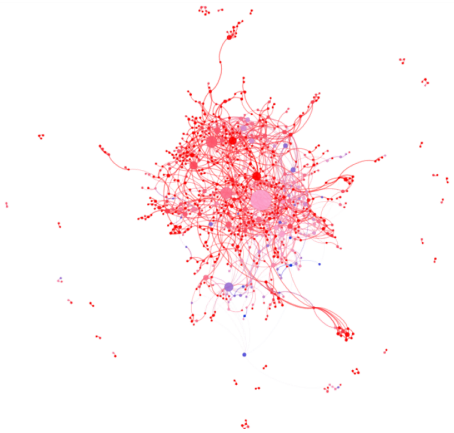

Fig. 3. Paper map of machine learning.

In paper maps, each node represents one paper in a research field, and each edge reflects the reference relation between two papers. If paper B is paper As reference, then theres an edge from A to B, and its directed. From the map, the node size shows a papers citation. And this citation count involves an aging algorithm. Due to the fact that old papers are more likely to have more citations than new papers. So if we just use the citation count to evaluate papers, this is unfair. Thats why we need to deploy an aging algorithm. The node color shows a papers age, newer papers with darker red and older papers with darker blue. Each edge has a weight reflecting the influence of one paper to another. Look at the whole map, through the layout, we may observe these results. The paper at the center may be a fundamental and important one in this research field. Maybe theres a new research trend in this field or some particular topics are hot in recent years.

C. Author Maps

Each affiliations have an author map. Here I choose Shanghai Jiao Tong University as an example shown in Fig. 4.

Fig. 4. Author map of Shanghai Jiao Tong University.

In author maps, each node represents an author of an affiliation. Each edge reflects the coauthor relationship between two authors, so its undirected. I use built-in PageRank algorithm to get the node size. Similarly, node color is gotten by using built-in partition algorithm. Edge weight is the strength of the relationship between two authors. Simply, it counts the times of being coauthors. Look at the whole map, from the layout, we may observe these results. The color partition may represents different research groups in this affiliation. The size of one partition to some extent shows this groups influence in the affiliation and also shows the collaboration relationship between those authors.

IV. Play with Webpages

Since I have generated all those maps, the last thing to be done is to put them online. Fig. 5 is a snapshot of author map webpage of Shanghai Jiao Tong University. Notice that users will not get exactly the same look of the webpage as this snapshot, because some displaying areas are user interacting and hidden by default.

At the upper-left corner of this webpage, it shows the map type, author map, and the name of this affiliation. When the mouse entering this area, more information about this affiliation will be displayed below. Currently it shows the total number of papers and authors in this affiliation. Users can also click the affiliation name and jump to the affiliation homepage on AceMap website. See also two small icons at the bottom-left corner, which is linked to two global maps.

In the middle, it is the map itself, also with some user interactive functions. When the mouse entering one specific node, this node together with its label will be highlighted. All related nodes and their labels will be highlighted too. Meanwhile, the detail information about this paper will be displayed in a floating window at the right side of the webpage. By clicking the node, users can directly go to the homepage of that paper and find out more. And there is a js function at

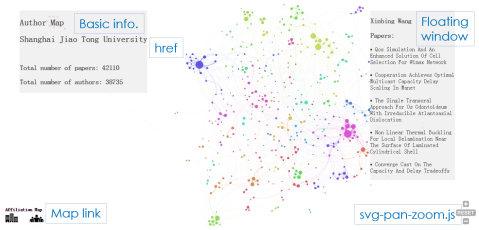

Fig. 5. Snapshot of author map webpage of Shanghai Jiao Tong University.

the bottom-right corner to enable zooming in and out on the map.

V. Future Work

As listed, there are few aspects of map feature of AceMap to be improved in future:

- Improve the look
  - Color
  - Size
  - Layout
  - Partition
- Show more information
  - Map
  - Floating window
- Possible user interaction
- Awesome global map

Last but not least, users' feedbacks and fellows' suggestions are the best guide for further improvements.

VI. Conclusion

Through this course project, I am able to master tools including SQL, Java, HTML language. I also learned the whole procedure to generate maps from data entries in database and finally put the visualization result online. This project helps AceMap website to have its unique feature: viewing academic big data from maps. Currently there are more than twenty-two thousand paper maps and over twenty-two hundred author maps online. Users can interact with the maps and jump to any other maps by the links. More features and functions are coming up on the website.

Acknowledgment

I would like to thank all the fellows in AceMap group for building the website. Also my group leader Guo Yunqi has given me plenty of useful suggestion on data processing and web front end techniques. At the visualization part, I got a lot of help from Lin Zebing, who showed me the framework of manipulating Gephi in Java. Their helps made it a lot easier for me to accomplish this project.