Academic Network Analysis and Hot Research Topic Modeling: A Google Scholar-based Approach

Chengdong Sun, Liang Wang, Shijuan Lin, Xin Tan
Harbin Institution of Technology, Shenzhen, China
SWS2023 Web Mining Group 1

Abstract

Google Scholar is a vital web search engine for academic articles, covering a wide range of scholarly literature including journal papers, theses, books, preprints, abstracts, and technical reports. It is an indispensable tool for the majority of researchers. However, it lacks features like displaying scholars' networks and predicting research trends. Our project aims to fill these gaps by using Python web scraping to create relationship networks, word clouds for research hotspots, and trend graphs. This will enhance Google Scholar's academic rigor and provide researchers with valuable insights into collaborations, recent interests, and emerging research areas.

Fig. 1: The icon of Google Scholar

Web Scraping

We utilized Python web scraping libraries, including **Selenium**, **BeautifulSoup**, and **Requests**, to extract information for our study. For scholars, we gathered data on their co-authors, citation counts, and homepages. Regarding academic papers, we retrieved information on citation counts, abstracts, PDF links, and webpage links. This data will be used for subsequent analysis and modeling.

Academic Network

We primarily constructed an academic network centered around He Kaiming, given his remarkable contributions to the field of Computer Vision (CV). As illustrated in Figure 2, we present an overview of his academic network.

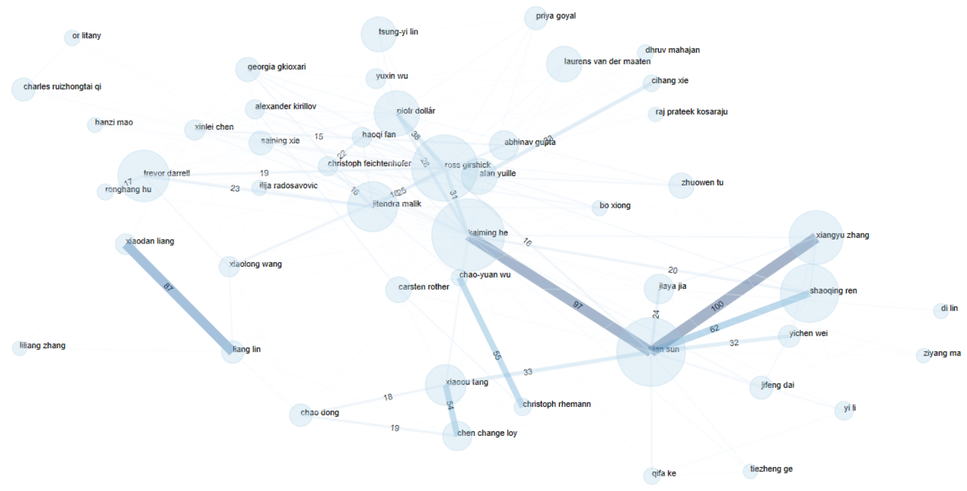

Fig. 2: An academic network

The size of the circles in the figure represents the respective scholar's citation count, while the thickness of the connecting lines between scholars indicates the frequency of their collaborative efforts. This academic network serves as an interactive and highly intuitive recommendation system.

Interactive Recommendation System

This academic network is presented as an HTML webpage, enabling users to navigate individual scholars' pages. Hovering over a scholar's circle displays pertinent details, including research interests and total citation count. This feature aids users in selecting more relevant scholars by visually showcasing collaborative frequency and citation count.

Academic Hotspots

Our analysis focuses on all papers presented at the CHI Conference on Human Factors in Computing Systems from 2016 to 2023. We aim to present specific field-oriented academic trends through the following forms:
The first representation is through pie charts, as Figure 3 shows the academic trends at the CHI Conference in 2023.

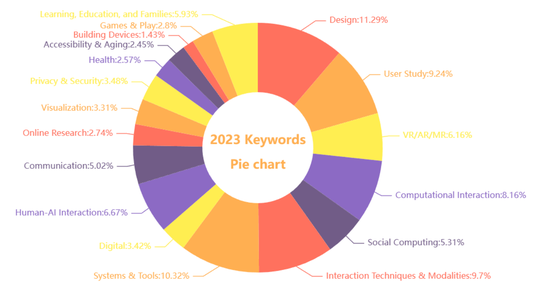

Fig. 3: The academic trends of CHI Conference in 2023

The second representation is through line graphs, as presented in Figure 4, to depict the changes in academic trends at the CHI Conference from 2016 to 2023.

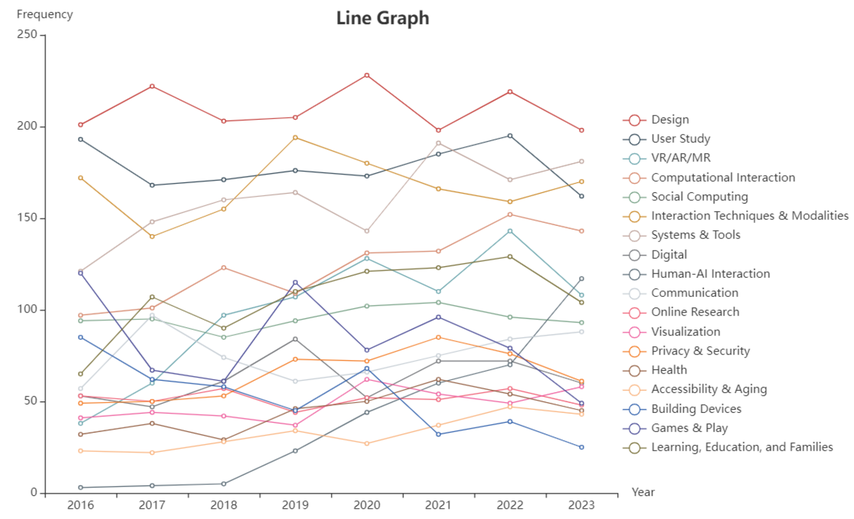

Fig. 4: Changes in academic trends at the CHI Conference from 2016 to 2023

Analysis and Conclusions

Based on the analysis of Figure 4 and taking practical considerations into account, the following conclusions can be drawn:

1. **Design** remains the most prominent direction in human-computer interaction, consistently having the highest number of research papers.
2. The overall trend for **VR/AR/MR** has been on the rise from 2016 to 2022 but experienced a sudden decline in 2023.
3. Both **Building Devices** and **Games & Play** directions exhibit clear downward trends.
4. **Human-AI Interaction** holds significant potential as artificial intelligence and large models become more prevalent, making it an important aspect of human-computer interaction

We employed the **ARIMA** model to forecast academic trends for 2024-2026, based on data from 2016 to 2023. The predicted trends align with expectations, as shown in Figure 5 (using Human-AI Interaction as an example).

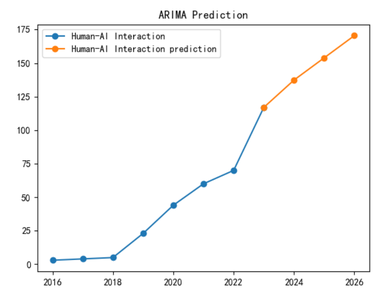

Fig. 5: 2024-2026 Human-AI Interaction prediction