College of Information Science and Technology

Analyzing the Impact of Sloan Digital Sky Survey on Astronomical Literature: A Multiple Perspective Approach

Jian Zhang*, Weizhong Zhu*, Yunan Chen*, Michael Vogeley and Chaomei Chen***
***College of Information Science and Technology, Drexel University, Philadelphia, PA 19104**
****Department of Physics, Drexel University, Philadelphia , PA 19104**

Introduction

A significant challenge in astronomy is posed by the enormous volume of new astronomical data being collected by sky surveys such as the Sloan Digital Sky Survey (SDSS) and the concurrent accelerated growth of the astronomical literature. This poster describes an ongoing and ambitious project[1] to support scientific discoveries in astronomy, in particular based on the use of SDSS data.

Our ultimate goal is to enable astronomers to explore and analyze across astronomical data and the corresponding literature. In order to achieve this goal, we first address a number of practical issues concerning the analysis of the impact of such large-scale survey datasets on scientific discoveries in terms of trends and patterns in scientific publications that utilize the data. In this work, we try to answer the following three questions.

- What is the basic trend of research literature that used SDSS data?
- What are the topics of those literature and their research frontier?
- In what context were those topics studied?

Data Collection and Analysis

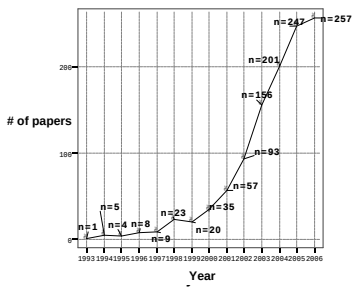
The SDSS research papers were collected from the *Web of Science*, retrieved with search terms "SDSS" AND "Sloan Digita*" over a time span between 1990 and 2006. The dataset contains a total 1,350 bibliographic records from 159 journals. We checked the abstract and title of each of the 1350 publications and removed irrelevant ones from the dataset. We removed three types of publications:

- We excluded 82 non-research papers, like review papers, letters, editorials, etc.
- We excluded papers had SDSS, which did not mean the Sloan survey, like Strategy Decision Support System, Spatial Decision Support System, etc.
- We excluded SDSS data-release papers and SDSS technical papers. Publications of these two types of papers are required to be in *Astronomical Journal* and citations to these papers are mandatory.

We take a multi-perspective approach, namely a statistical perspective, a citation analysis perspective, and a text analysis for each of the questions we asked perspective. Also some new developed visualization software, like CiteSpace (Chen, 2006) and Storylines (Zhu & Chen, 2007), were used.

What is the basic situation of research literatures that are based on SDSS project?

Annual publication rate



Top 5 productive authors

Rank	Author	Publication	Citation
1	Brinkmann J	242	9280
2	Schneider DP	211	7669
3	York DG	125	6266
4	Ivezic Z	118	6495
5	Strauss MA	114	5880

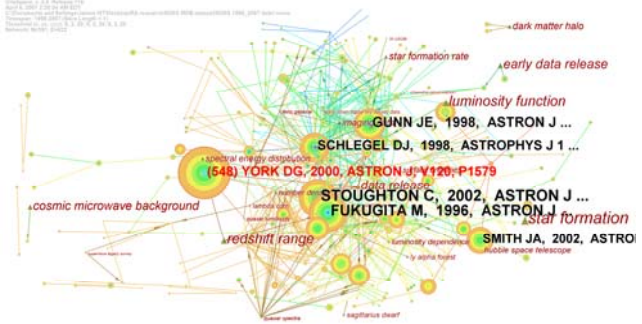
Top 5 productive institutions

Rank	Institution	Publication	Citation
1	Apache Point Observatory	251	9315
2	Princeton University Observatory	250	9869
3	Johns Hopkins University	226	8819
4	Penn State University	221	7708
5	University of Chicago	200	8016

What are the topics of those literatures and their research frontier?

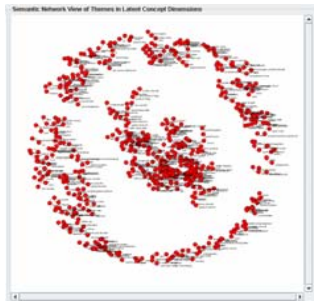
Timeline of topics (burst terms) from 1998 to 2006, created by *CiteSpace*

Visualization of co-cited papers and topics (burst terms), created by *CiteSpace*

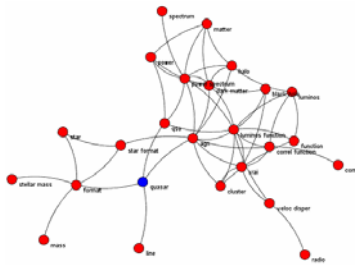


In what context were those topics studied?

Concept mapping of SDSS in 530 dimensions, created by *Storylines*



Major concepts on latent dimension 1, created by *Storylines*



Acknowledgements

This work is in part supported by NSF grant (No. IIS-0612129). Thomson ISI provides part of the bibliographic data for the analysis

References

[1] Detailed information and preliminary results about the project are available on http://cluster.cis.drexel.edu/~cchen/projects/sdss/

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*. 57, 359-377.

Zhu, W., Chen, C. (2007) Storylines: Visual exploration and analysis in latent semantic spaces. *Computers & Graphics*, 31, Special Issue on Visual Analytics.