Information Visualization Research: Citation and Co-Citation Highlights

Chaomei Chen[*]
Drexel University

Abstract

An overview of the entry is given. The techniques used to prepare the InfoVis contest entry are outlined. The strengths and weaknesses are briefly discussed.

Keywords: citation analysis, co-citation analysis, domain visualization

1 Overview

The entry addresses the first task with three different visualizations generated by a recent Java visualization application CiteSpace [1] and an earlier VRML scene generator developed under the GSA framework [2-4]; both tools have been developed by the author.

Both tools extract citation and co-citation patterns from the contest dataset and present the patterns in a three-dimensional model in VRML and a two-dimensional model in CiteSpace. A citation is an instance of the reference to an article. A co-citation involves citations of two papers together in subsequent publications. Co-citations have been used as a clustering mechanism in citation analysis.

In order to identify emergent themes in the history of the literature, terms are extracted from title, abstract, and keyword list of citing articles in CiteSpace. Sharp increases of popularity of terms are identified as a sign of a surge of interest or the birth of a new thematic thread. Surged terms are embedded in a typical co-citation network of articles as indicators of such themes. CiteSpace supports two related views: a timezone view and a traditional node-and-link view. In the timezone view, the network is superimposed on a time zone grid so that items of a given year are restricted to corresponding time zones only. It is particularly useful for tasks such as identifying the chronological sequence of thematic surges. The node-and-link view, on the other hand, uses widely known spring embedder layout. It is useful if the network consists of a number of distinct clusters. The citation tree-ring feature provides an intuitive way to identify highly cited papers – papers that have large tree-rings. The color pattern of a tree-ring can also tell us the citation distribution of the underlying paper. CiteSpace provides a variety of analysis facilities for users to analyze a dataset in several ways. For example, CiteSpace allows the analyst to divide the entire time interval into time slices and merge individual networks.

The major advantage of the VRML models is their animation feature, which can vividly replay the history of an evolving field. The complex chronological relationships among a large amount of papers can be relatively straightforward to understand in the animated visualization. Questions such as which areas appeared first in the field can be easily answered by such animations. Detailed descriptions of the two systems can be found in the references.

----
[*]e-mail: chaomei.chen@cis.drexel.edu

2 Tasks

Three visualizations are provided for task 1 – the overview task. The first one is a 3-dimensional VRML scene, the second one is a network view generated in a typical graph drawing manner in CiteSpace, and the third one is a timezone view by CiteSpace as well.

The most obvious insight is that the field has two prominent pioneer papers: the cone tree paper by Robertson et al. and the fisheye view paper by Furnas. The first overview shows that InfoVis symposium papers tend to appear in the second specialty identified by Principal Component Analysis (PCA) of co-citation patterns, whereas the largest specialty appears to be formed by papers outside the symposium, especially papers published prior to the beginning of the symposium series. The second and the third views from CiteSpace demonstrate the advantages of including surged terms, which helps us to identify the nature of emergent themes in the history and papers that were associated with these themes.

Two visualizations are provided for task 2 – the evolution task. The first one is a 3-dimensional animated visualization, chronologically showing the growth of the bibliographic network in accordance to the sequence of publication and citation events (See Figure 1). The timezone view identifies various themes.
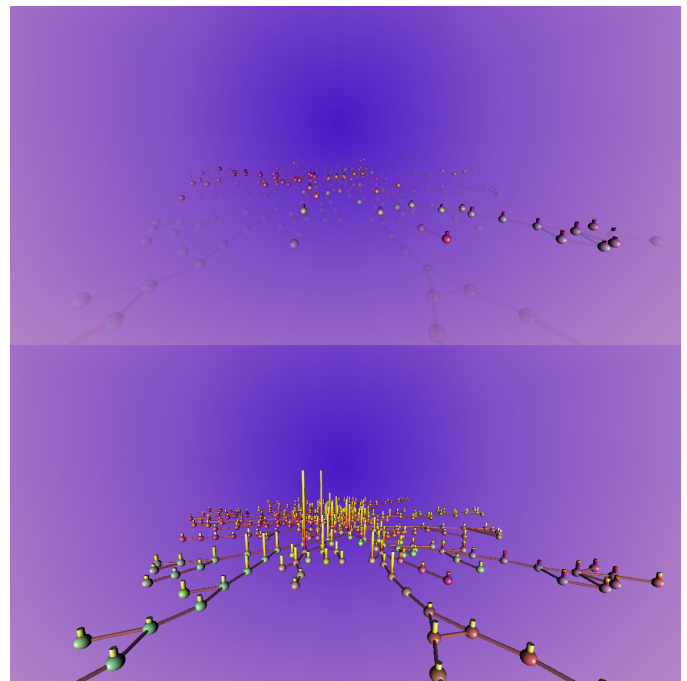

Figure 1. Animated visualization of the growth of the field.

Figure 2 shows a TimeZone view. Surged terms trailblaze emerging themes across time zones, followed by articles that were referenced in association with such themes, i.e. by terms-to-article citations.
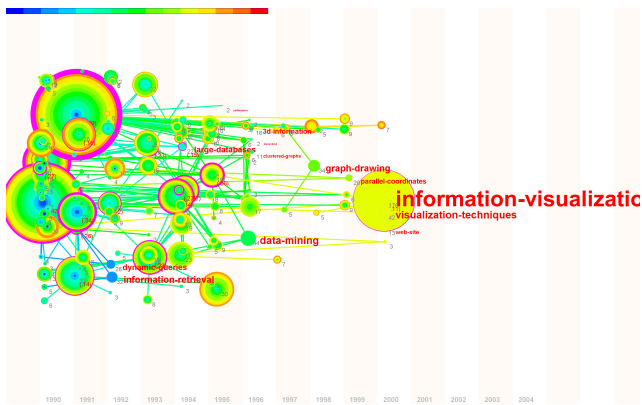
Figure 2. Surged terms pointing to the key references associated with thematic trends.

Tasks 3 and 4 are briefly addressed. An interesting observation is the absence of distinct clusters. In our experience, we typically observe a number of clusters in a field. In part, the lack of clusters may be due to the nature of the dataset, i.e. all citers are limited to one single symposium series. If other sources such as graph drawing were included, it is likely that more clusters may emerge.

Note

I include all images in the same standard form because the submission webpage didn't seem to allow multiple entry submissions. Only one video clip is in the submission pack, as it appears that the submission webpage has overwritten previous ones. Additional video clips are available at:
http://www.pages.drexel.edu/iv04contest/citespace-gui.avi
http://www.pages.drexel.edu/iv04contest/citespace-display.avi

A video file of the VRML scene is 65Mb after compression. I'd be happy to supply the video via other means upon request (e.g. put it on a CDROM).

References

[1] Chen, C. 2004. Searching for intellectual turning points: Progressive Knowledge Domain Visualization. In *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 101, Suppl. 1, 5303-5310.

[2] Chen, C., AND Morris, S. 2003. Visualizing evolving networks: Minimum spanning trees versus Pathfinder networks. In *Proceedings of the IEEE Symposium on Information Visualization* (InfoVis'03). Oct 19-24, 2003. Seattle, Washington. IEEE Computer Society. pp. 67-74.

[3] Chen, C., AND Kuljis, J. 2003. The rising landscape: A visual exploration of superstring revolutions in physics. *Journal of the American Society for Information Science and Technology* 54, 5, 435-446.

[4] Chen, C., AND Paul, R. J. 2001. Visualizing a knowledge domain's intellectual structure. *IEEE Computer* 34, 3, 65-71.