CPSC533c Information Visualization

Instructor: Dr. Tamara Munzner

Term Project Paper

Analysis of InfoVis 2004 Contest Data:

A Survey & Analysis

Chia-Ning Chiang
Ron Jung-Rung Han
April 21, 2004

1. **Description of problems**

The task domain is to visualize the field of Information Visualization. This is the InfoVis 2004 contest problem. The contest is a programming project, where contestants design their own visualization programs to effectively produce a visualization. However, in this project, we are doing a survey project: to evaluate existing visualization tools, in particular, VxInsight, on their capabilities to explore the dataset and perform the InfoVis contest tasks.

We use the dataset provided in the InfoVis 2004 contest. The dataset contains metadata for all papers of 8 years (1995-2002) of InfoVis conference and their references. The metadata was largely taken from the ACM Digital Library. It includes publication title, authors, keywords, abstract, references, and etc.

The contest proposes several tasks to be solved with visualizations. Taken directly from the InfoVis Contest website, they are the followings:

(1)   Create a static (non-interactive) overview of the 10 years of the InfoVis conference.

(2)   Characterize the research areas and their evolution.

(3)   Where does a particular author/researcher fit within the research areas defined in task 2?

(4)   What, if any, are the relationships between two or more or all researchers?

We will use several light-weight tools to explore the possibilities of each of them answering the above task questions. Then we will focus on VxInsight, a commercial grade visualization tool developed by the Sandia National Laboratories, to do these contest tasks.

2.   **Related Works**

Visualizing Science by Citation Mapping

Visualizing domain specific researches have been undertaken for many years. The majority of these studies are performed at the discipline (White & McCain, 1998; Spasser, 1997; Boyack et al, 2001b) or specialty level (Boerner, 2000, Boyack, 2000). Researchers use advanced bibliometric[1] measures and indicators to reveal the development of scientific knowledge. Application areas include studies of scholarly communities and networks, the

growth and evolution of fields, the diffusion of research topics, and the relationships among individual authors and institutions. (Boerner et al, 2003)

Science and technology maps are most often based on computed similarities between journal articles using citation analysis (Small, 1999), or cooccurrence or coclassification using keywords, topics, or classification schemes (Nederhof & van Wijk, 1997; Noyons & van Raan, 1998; Spasser, 1997). The number of publications produced by a researcher or group over time is the simplest indicator available. Citation counts allow one to find out where and how often an article is cited, which provides an estimation of the importance of an article. Studies to identify intellectual or social networks are performed using author cocitation[2] analysis (Chen et al, 2001a) or on the basis of coauthorship (Newman, 2001). Co-occurring words or citation links among documents, authors, terms, or journals, and have been described in detail in Boerner (Boerner et al, 2003), White & McCain (White & McCain, 1997), and White (White, 2003).   Other approaches include linguistic, content, or semantic similarities (Landauer et al, 1998; Boerner, 2000) as computing resources have increased and information visualization techniques have progressed to enable the use of geometric proximity in the graphical user interface to capture relationships between objects.

Once relationships between objects (articles, terms, authors, etc.) have been defined and a similarity matrix (based on cocitation or cooccurrence, etc.) has been computed, algorithms are used to cluster the data. Common clustering methods for producing maps include hierarchical clustering, k-means algorithms, multidimensional scaling, principal components analysis, and self-organizing maps (Boycek, 2003). Traditionally, the standard output for literature mapping studies has been a circle plot where each cluster was represented by an appropriately sized circle. Links between circles provide relationship information. In recent years, several systems have been reported that use a computer display to produce the mapping and allow some navigation of the map space (Boyack et al, 2001a).

---

[1] Bibliometrics Bibliometrics utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature.
[2] Co-citation: to cluster cited articles according to their co-occurrences in journal article reference lists.

Visualization Tools

VR-VIBE uses statistical techniques to visualize a document bibliography and allows the user to interact with and manipulate the space. The VR-VIBE environment is constructed from a bibliography file and a corresponding set of queries. The queries are sets of keywords which can be searched for within the content of the documents in order to generate scores to show the relative correlation between the query and the document (Benfor, 1995). The Cat-a-Cone integrates search and browsing of very large category hierarchies with their associated text collections. One key insight is the separation of the representation of category labels from documents, which allows the display of multiple categories per document. Another key component is the display of multiple selected categories simultaneously, complete with their hierarchical context (Hearst & Karadi, 1997a; Hearst & Karadi, 1997b). SCI-Map developed by ISI (Small, 1999). SCI-Map uses a hierarchically-nested set of maps to display the document space at varying levels of detail. This nesting of maps allows drilling down to subsequent levels. Each map is similar to the traditional circle plot, where the size of the circle can indicate the density of documents contained in the circle, or some measure of importance. Relationships at each discrete level are indicated by links between circles. The SPIRE suite of tools, which originated at Pacific Northwest national Laboratory (Hetzler, 1998; Wise, 1999), map objects to a two-dimensional plane so that related objects are near each other, and provides tools to interact with the data. SPIRE is exclusively focused on textual objects and computes similarities using text analysis. SPIRE has two visualization approaches: a scatter plot of the text documents and a static terrain display. SPIRE is very similar to VxInsight, except SPIRE does not provide the continuous, multi-resolution viewing capability and lacks the flexible database interface.

Pathfinder Networks (PFNETs) is an approach taken by White (White, 2003) for visualizing author cocitation analysis. In PFNETs, nodes represent authors, and explicit links represent weighted paths (e.g. cocitation counts) between nodes. White's paper directed us to test on the AuthorLink[3] system. AuthorLink is a query-based visual interface at Drexel University. Buzydlowski and coworkers have developed an interface to over 1.2 million records from

---

[3] AuthorLink url: http://cite.cis.drexel.edu

the Arts & Humanities Citation Index (AHCI). The user types in the name of an author of interest, and a map of the 25 authors most linked to the query author is returned. The user can drill down through an author to find individual works. .

A four-step procedure for visualizing intellectual structures, developed by Chen (Chen & Paul, 2001b) that extends the transforms traditional author co-citation analysis by extracting structural patterns from the scientific literature and representing them in a 3D knowledge landscape. Users can apply such visualizations to discover patterns and make valuable connections between data.

VxInsight is the one we can have a hands-on trial. VxInsight was developed to build and explore maps of technology using data from the Science Citation Index (SCI) (Davidson, 1998). VxInsight allows users to navigate through 3D landscape using mouse to zoom, pan, and rotate. It displays fields from individual records. Labels peaks are dynamically sampling words in fields. VxInsight displays trends over time. VxInsight has been explored to a broad application to mapping and navigation of many different types of data (Boyack et al, 2000; Boyack, 2001; Boyack et al, 2002; Boyack et al, 2003; Werner-Washburne et al, 2002).

3.   Exploration of Tools -- Before Using VxInsight

Prior to having a trial on VxInsight, we felt that we need to understand how to measure bibliographic coupling and co-citation analysis besides citation counts, we had tested on a free ware called Bibexcel.[4] Bibexce is a free-ware, developed by Professor Olle Persson, Infosk, Umea univ, Sweden. This tool-box is designed to assist a user in analyzing bibliographic data. This tool-box includes a number of tools, some of them visible in the window and others hide behind the menus. It takes several data formats from indexes and abstracts databases, such as ISI's Web of Science, Medlin, Biological Abstracts, and Dialog Online Information Services, etc., and processes and builds them into tab delimited table; in addition, it also has build-in calculation and fractionalization functions to generate the counts and coordination required to generate the citation map. For file types see Appendix A. It is a step-by-step tool of preparing data for visualization.

---

[4] Bibexcel url: www.umu.se/inforsk/Bibexcel/

However, Bibexcel needs a Multi-dimensional Scaling-program (MDS) to be able to produce maps, and the MDS-software is not distributed together with Bibexcel. MDS attempts to find the structure in a set of proximity measures between objects. MDS has been one of the most widely used mapping techniques in information science, especially for document visualization, author co-citation analyses, science mapping, and visualizing group memories, and performance assessment, etc. (Boerner, 2003). Biblexcel is designed to let Systat/Dos version 4.1, which we don't have, to handle of the scaling. Even though we cannot generate the citation maps, we did learn about the relationships among various data files and the idea of bibliometrics. Through our exchanges of emails, Professor Persson pointed us to find a citation map from his website, which can demonstrate Bates' author co-citation map (Figure 1).
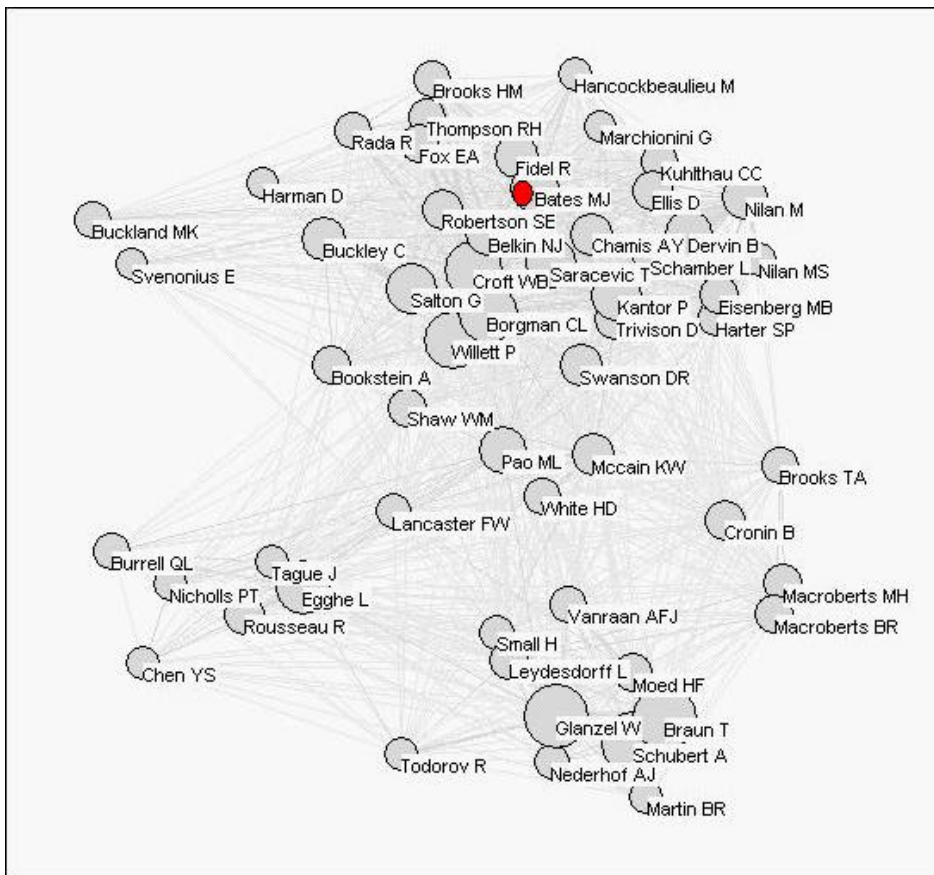

Figure 1. Influential Authors in Library and Information Science 1986-1996.
    An All Author Co-citation Map, by Olle Persson. (Thanks to Professor Persson)
    http://www.umu.se/inforsk/LIS/index.h

To understand how the Pathfinder Networks works, we logged on to AuthorLink, a web-based system developed at the Drexel University that creates live interfaces for cocited author retrieval on the fly. With the input "Bates", the system generates three different maps (Figure2, Figure3, and Figure4). Based on this hands-on experience and a few papers that compare Pathfinder Networks and Multidimensional Scaling, we believe that the Pathfinder Networks is currently the superior tool to show the Author co-citation Map.
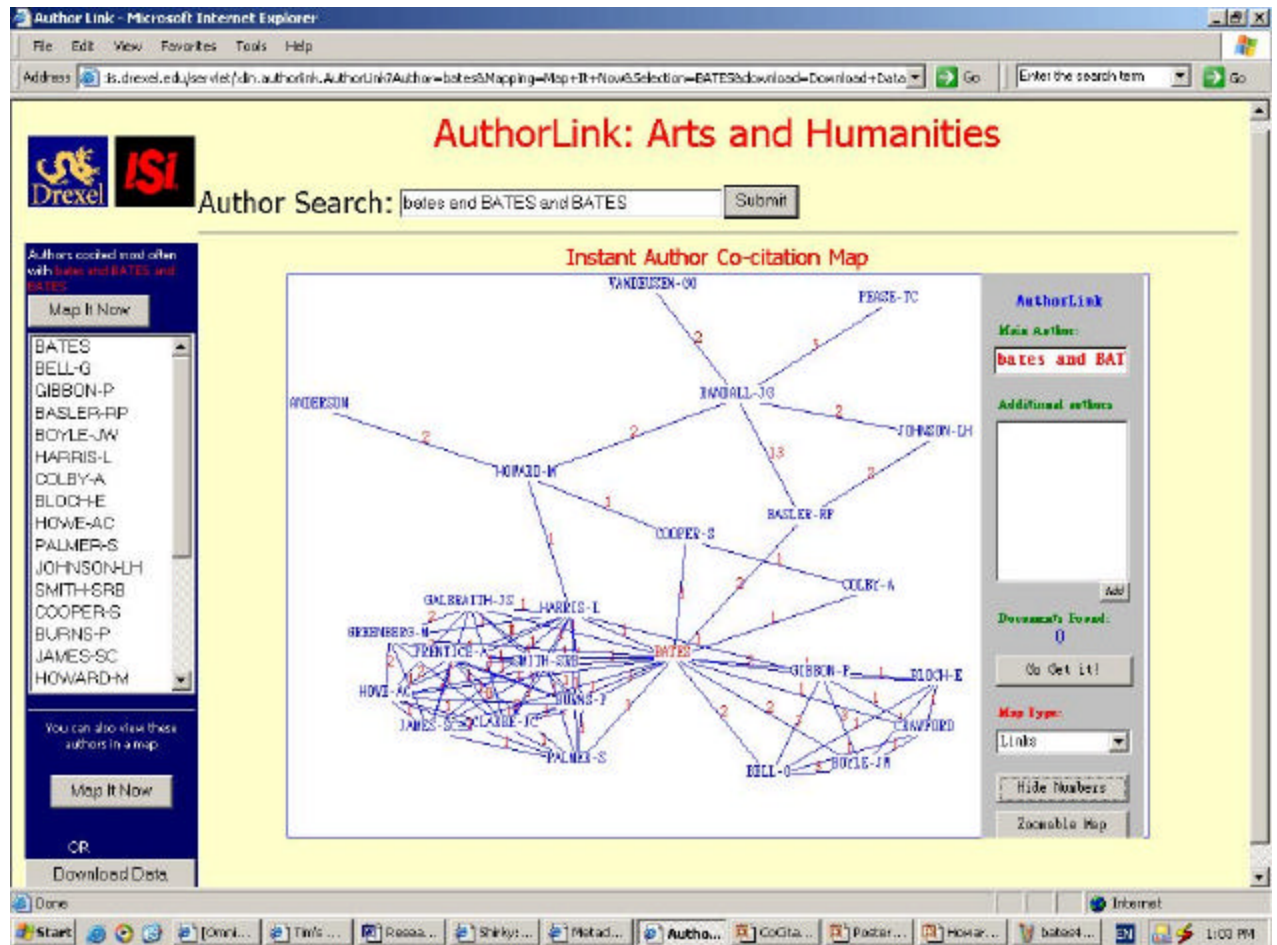

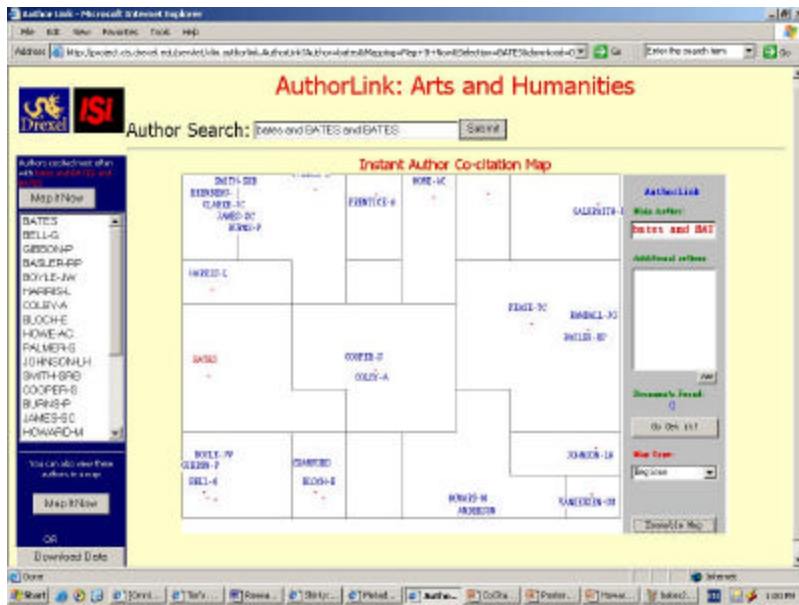Figure 2. Instant Author Co-citation Map of Bates in a hierarchical map.

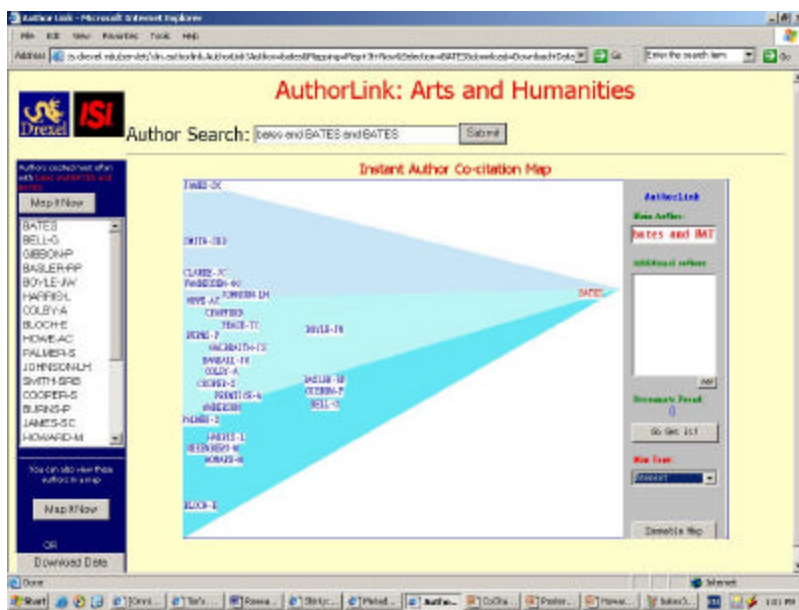Figure 3. Instant Author Co-citation map of Bates in a tree-map

Figure 4. Instant Author Co-citation map of Bates in a triangular-shaped map

4. Setup VxInsight

There are several steps required to display our dataset in VxInsight as a 3D landscape. In high level terms, these steps are the followings:

- Provide several files required by VxInsight. The most noteworthy one is the input dataset that is put in an ODBC-compliant database.
- Specify a similarity measure between data objects.

- Generate [x,y] positions for each data object from the similarity measure. This step, known as ordination, is done by VxOrd (a subsystem of VxInsight). A similarity measure is optional.
- Load the data into VxInsight.

We wrote a parser program to convert the contest data, in XML format, into several tab delimited text files, each of which represents a table in a relational database. Then we load the tables into Microsoft Access to create the database file. Other files that VxInsight needs include a database alias file which contains a reference to the database file; a VxInsight configuration file; and a "stopword" file which is used to prune the words used to display labels.

Then we create a table containing two columns: first column is a paper ID and the second column is paper ID of a paper that is cited by the paper in the first column. We could specify a similarity in a third column, however, if omitted, VxOrd will automatically use co-citation as a similarity measure.

The next step is to load the database file into VxInsight, then load the coordination file generated by VxOrd to place each data nodes in a meaningful way.

5. Exploration of the dataset

We began our analysis by scaling down the year range once we started the VxInsight image. We can see research areas and their evolution as the following: on a full range from 2004 to 1974, the top of the terrain showed four major areas generated from common words in the title fields. As we move the slider back one year at a time, we can observe the trends. In Table 1., we can see the evolution shift in 2001, 1995, 1994 and 1993.

2004	Visualization/data	Information visualization	Visualization/data	Information Visualization	Algorithm/animation
2001	Visualization/ information	Information Visualization	Visualization/graphs	Information Visualization	Algorithm/animation
1995	Visualization/ Structure	Information/ Graphical	Visualization/data	Information/ Dynamic	Parallel/ Visualization
1994	Structures/ Approach	Information/ Graphical	Data/Visualization	Information/ Dynamic	Algorithm/animation
1993	Approach/ Interface	Information/ edition	Rendering/ Volume	3D/ Visualization	Algorithm/animation

Table 1. Major co-terms from titles appeared on the screen that changes by sliding the time line

By sliding the year range, we found that the earliest paper (1974) is about vector space, 1975 about clustering algorithm, 1976 about hierarchical geometry, 1978 about ordination, 1979 data structure, until the early 1990s researches are shift from algorithm, data structure, to rendering data. From 1994 and up, the major research areas are more focused on five clusters show in Table 1.

Using VxInsight's query function, we found that the IEEE produced the most papers (241 out of 613), and the ACM is in the second place (68 out of 613). Querying the title terms is one way to find out how many papers have been published on a topic. We decided not to use common terms such as information, visualization, data, and graphical etc., in stead we tried terms of specific InfoVis techniques, e.g. pan (0 hit), zoom (8 hits), fisheye (7 hits), hyperbolic (6 hits), focus (7), focus+context (4), taxonomy (7), animat* (18), tree (30), and algorithm (36). Such a query returns a list of matched objects, which will be highlighted as color dots. By examining a particular cluster and turn on the links we can see the citation structure among the papers. A shift click on a particular object brings the metadata of the object on top of the screen.

ISI Web of Science and Citerseer Citations

To determine who the dominant people are, their institutions, and the type of their publications, we used the ISI's Web of Science and the CiteSeer to examine the citation context of most cited authors in the information visualization field. Although ISI contains less proceedings, but it has a broader coverage of disciplinary area; on the other hand, the CiteSeer is more focused in computer science and related researches. In the InfoVis dataset there are 613 articles that produced 4386 cited references over a period of 30 years, from 1974 to 2004 (Figure 5). We used the Pivot table function of Excel to find out the top 10 most cited references (Appendix B). The most cited articles comes from Xerox Palo Alto Research Center and University of Maryland. There are eight conference papers (seven ACM SIGCHI and one IEEE proceedings), one book (Tufte, 1986), and two journal articles published by ACM.

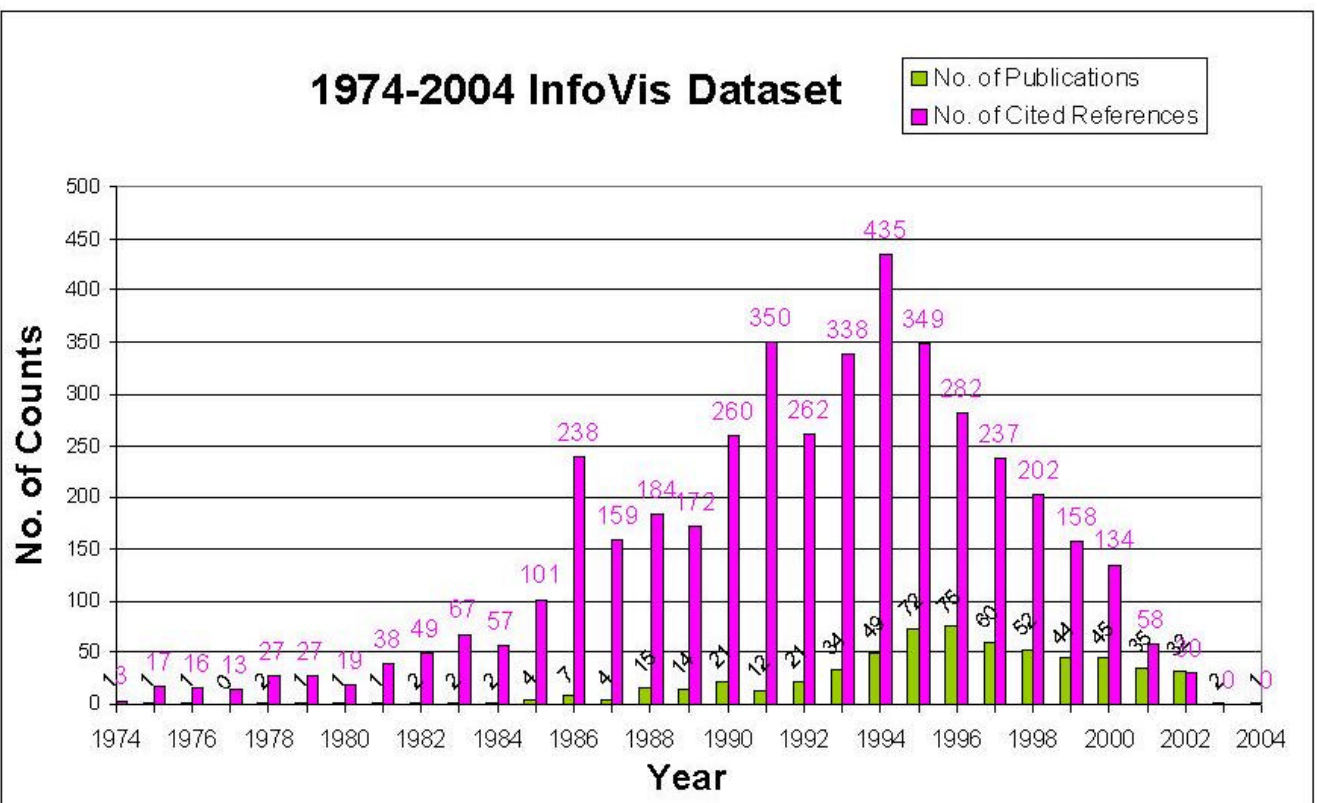| Year	No. of Cited References	No. of Publications
1974	3	1
1975	17	1
1976	16	1
1977	13	0
1978	27	2
1979	27	1
1980	19	1
1981	38	1
1982	49	2
1983	67	2
1984	57	2
1985	101	4
1986	238	7
1987	159	4
1988	184	15
1989	172	14
1990	260	21
1991	350	12
1992	262	21
1993	338	34
1994	435	49
1995	349	72
1996	282	75
1997	237	60
1998	202	52
1999	158	44
2000	134	45
2001	58	35
2002	30	32
2003	0	2
2004	0	1

Figure 5. A timeline of the no. of publications and the no. of citations from 1974 to 2004

The most cited article is "Cone Trees: animated 3D Visualizations of Hierarchical Information", which has been cited 65 times, by Stuart K. Card of Xerox Palo Alto Research Center. The ISI' s Web of Science (from 1965 to 2004) contains 68 citations and the Citeseer contains 64 citations. The searches found that there aren' t much difference. However, the Furnas ' s 1986 paper receives 61 as the second most cited author; we found 171 counts from ISI, which has surpassed the citation counts within the InfoVis field, and more interestingly, we found out that the CiteSeer has 237 citations to his paper. This may mean the "Fisheye views" has great impact on InfoVis researches. The "perspective wall" is the third highest, again by Mackinlay, Roberson, and Card of Palo Alto Research center, which has 66 in ISI and 134 citations in Citeseer. Tufte' s book, ranked the fourth, is the only book entering the top 10, which means it has a great impact on InfoVis research. Then we examined how many citations the University of Maryland has, we found that Ahlberg & Shneiderman found 76 citations in ISI and 153 citations in Citeseer; and Johnson & Shneiderman' s paper only found 9 citations in Citeseer, and none from the ISI' s Web of

Science search.   This citations context analysis showed us that it's obvious that "more authors attracts more citations" is not true here, e. g. the number of citations appeared high peaks at the year of 1986 and 1991, comparing to those two years the number of papers published.

To answer the question of where does a particular author/researcher fit within the research areas? We queried "Munzner" and got 6 hits, and those papers fall scattered among the mountains labeled "data/visualization" and "visualization/graphs" (Figure 6).   We think that unless we have further data, such as keywords or classification codes to better present the research areas, we cannot simply by using the title words to demonstrate this in a more appropriate way.
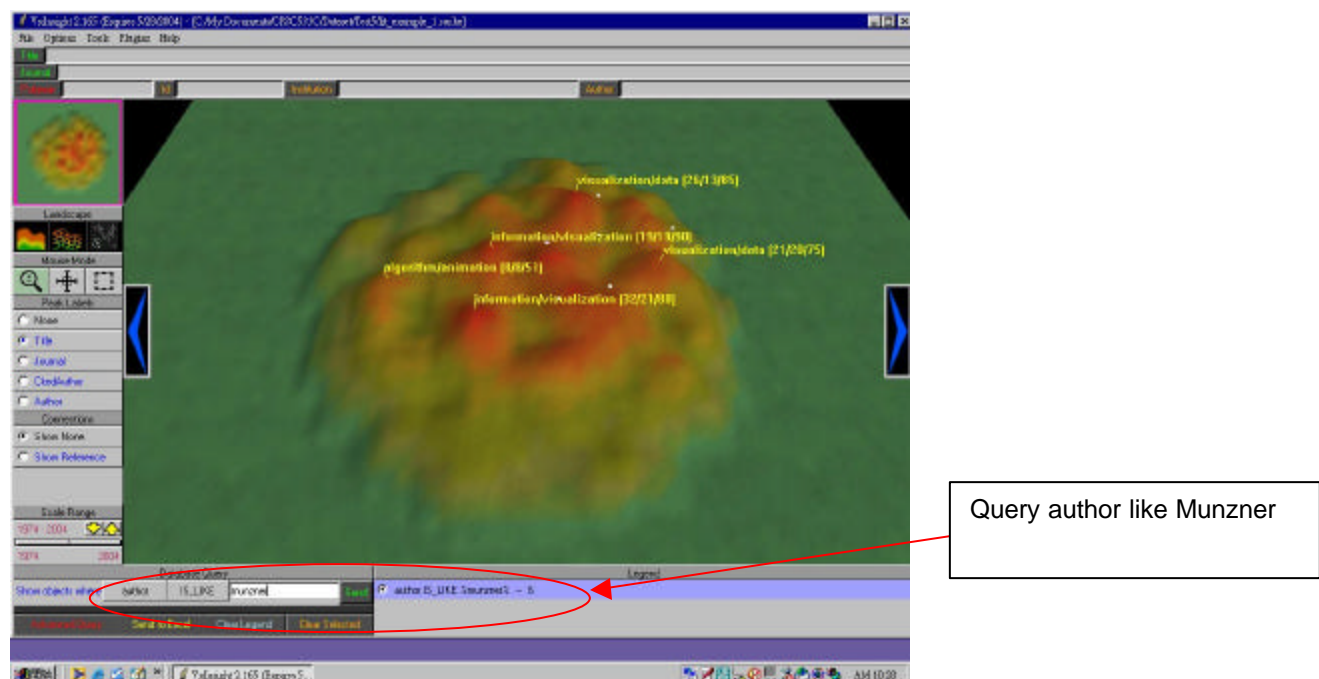

Figure 6. Use VxInsight Query function to search author like Munzner, and get 6 hits (white spots on the screen)

What are the relationships between two or more or all researchers?

From our hands-on testing of the Pathfinder Networks and the VxInsight, we both feel that the Pathfinder can give a much better author-cocitation relationship map. In Figure 6 shows the most cited authors – Stuart Card, Furnas, Tufte, and others; however, we cannot tell how they related to each other. On the other hand, the Pathfinder Networks can give a much clear sense (Figure 7 and Figure 8).
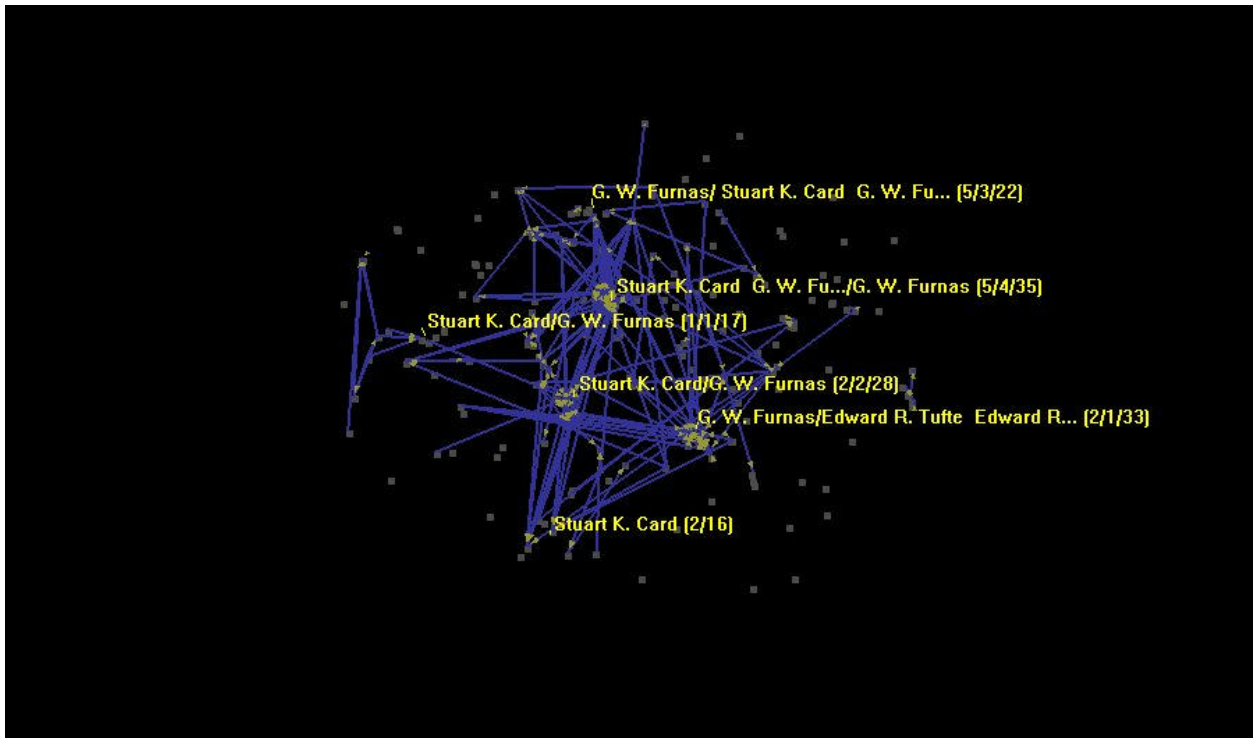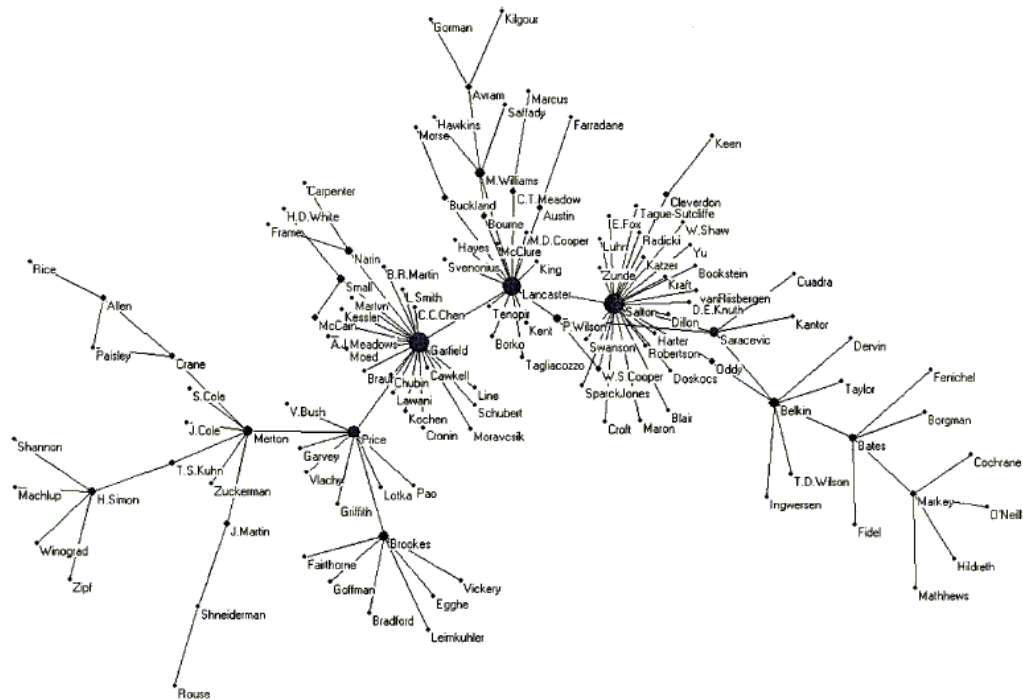
Figure 7. Author co-cited map from VxInsight

FIG. 1. PFNET of 121 information science authors based on raw cocitation counts.

Figure 8. This is a figure taken from White's paper that displays the relationships among authors based-on author co-citation analysis

What are the relationships between journal articles, conference papers and technical reports, and thesis? We feel that the InfoVis 2004 dataset is not large enough to answer this question. Because it contains only one thesis and few technical reports, it is required to collect more data from indexes & abstracts databases, such as Networked Computer Science Technical Reference Library (NCSTRL), The National Technical Information Service (NTIS), and UMI' s Dissertation Abstracts, etc.

For Additional dataset tasks, we did two more datasets. One is InfoVis class readings, the other is a dataset of 148 papers that cites Marcia J. Bates' 1989 "Berrypicking theory" paper from the ISI' s Web of Science.

For readings data set of InfoVis courses, our approach is to collect those information visualization papers posted on the course websites of Stanford University, the University of California at Berkeley, and the University of British Columbia. We then used VxInsight to produce the relationship among these three dataset. In Figure 10 showed that the readings of information visualization courses in these three universities are quite different.
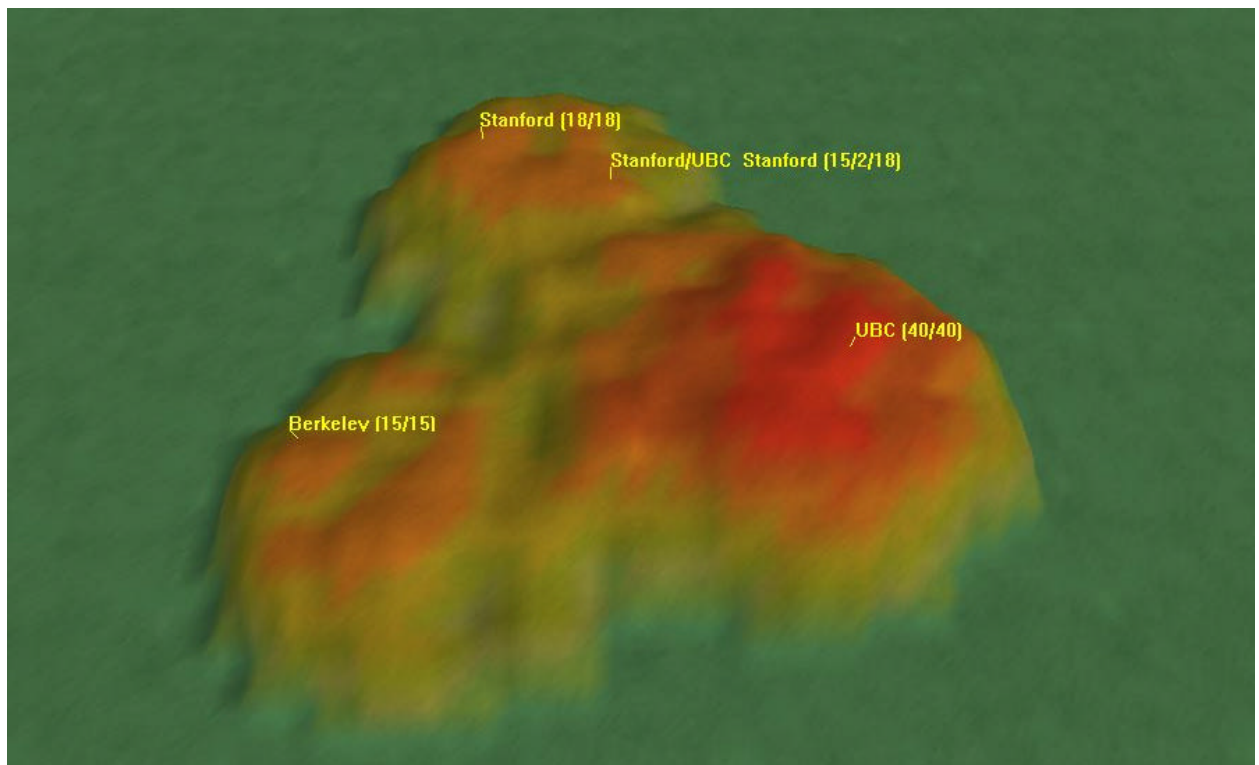

Figure 9. – InfoVis Course dataset shows the difference of topic areas in those universities.

We had a chance to demonstrate the VxInsight to a groups of faculty and students in library, archival, and information science in a poster session on Marcia J. Bates' "Berrypicking theory". To make the audience feel more attached to the topic we searched the ISI's Web of Science and collected all citations on Bates' 1989 paper. The dataset contains 148 papers and 8584 citations from 1990 to 2004, and we used it to build a VxInsight dataset, so that the audiences not only to see an advanced information visualization techniques, but they can actually have a hands-on experience. The usability result is reported in the user study section of this report. The co-term display of the bates dataset produced a much clearer research topics -- information/design, information/navigation, information retrieval, and electronic interaction, and information feedback, these research areas are exactly what Bates' Berrypicking theory falls, shown as Figure 11
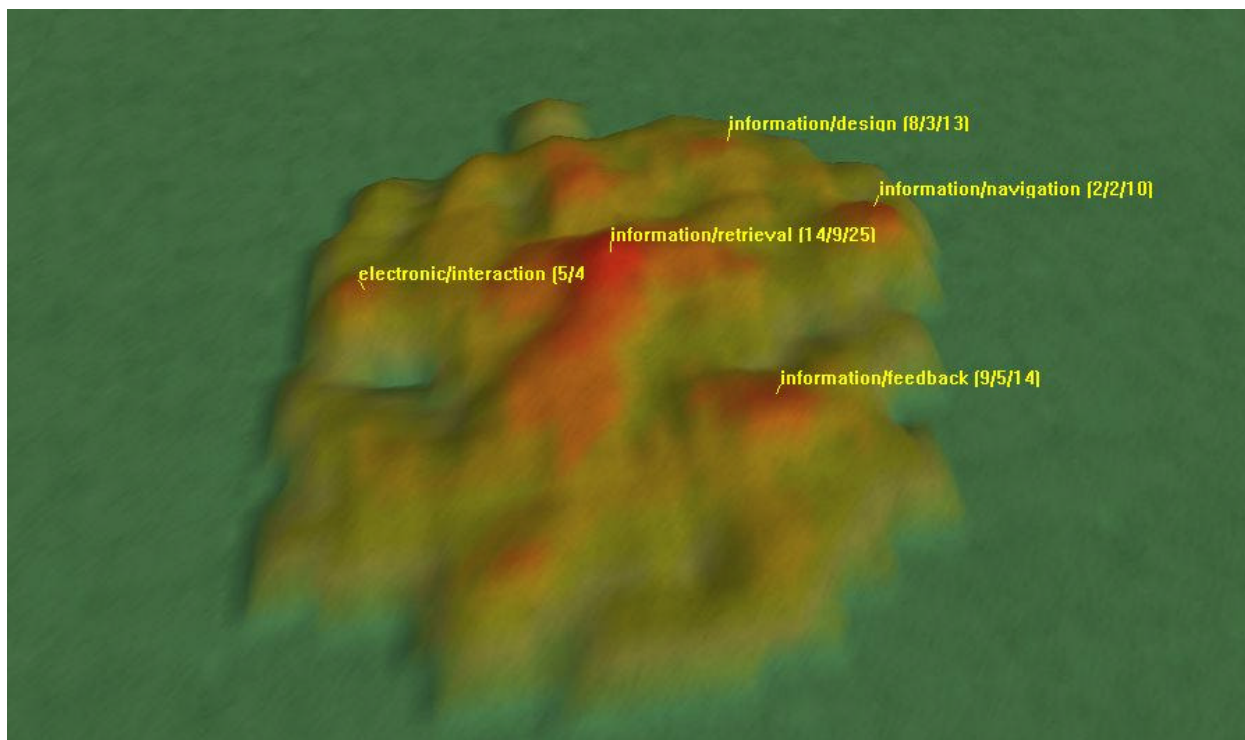

Figure 10. Bates dataset shown in VxInsight image. The labels are all clearly identify the research topics that related to Bates theory.

Although three different analyses are presented here, the process followed to conduct each study was the same. First, relevant and appropriate data were procured typically by query to a bibliographic database (a Microsoft Access database). Second, and object-to-object (e.g., article-to-reference or article-to-course_code) similarity was calculated. This step can require the use of database functions, statistics, or other mathematical processing depending

upon the type of similarity chosen. Third, the data were clustered using the VxOrd algorithm. Fourth, the data were loaded into VxInsight for viewing, navigation, discovery, and analysis.

6. Evaluations

(1) Ease of use

- Excel: Excel is the most handy tool for filtering out the data patterns we want.
- Bibexcel: tedious work in terms of ease of use. Can' t get the map without SPSS or SYSTAT/DOS, (one of the Multi-Dimensional System)
- AuthorLink: no way to manipulate the data, it online dynamically builds the graph.
- VxInsight: easy to install, build, navigate. It has the flexibility and dynamic pan and zoom function and the ability to show the element of time and evolution.

(2) VxInsight User study

We had a chance to do a user study by observation. We demonstrated VxInsight citation map to a group of faculty and students in library, archival, and information science.   There were 15 subjects approaching the system and expressed their interests to know about how the system works. Two faculty members did the trials by themselves. One faculty member (male, age 30-40, with computer science background) performed the test by himself. He had successfully clicked all the functions, except query, —zoom, rotate, scale range, show labels, change landscapes, select peak and send to Excel; however, he didn' t know how to zoom back when he zoomed in and lost the landscape on the screen. The other faculty member (female, age 50-60, with information science background) did the same tests like the other faulty member, and only failed to select peak and send to Excel. The other 13 subjects all asked us to perform the demonstration, among them only one interested to know more about VxInsight and asked for a copy of VxInsight' s website printouts from us. In general, their opinions are as follows:

- They all expressed how incredible of this tool.
- Two faculty members, who did the trials by themselves, expressed the intuitive of navigation of VxInsight.

- They first didn't get the idea of the terrain landscape, after we had explained the metaphor, and they all expressed this is a good idea.
- They liked the scale range idea, but they didn't like the name "scale range"; and they would like to see the most recent year range, for example, they asked "Can you show the range between 2002-2004, rather then all the way back to the earliest date?" Because in bibliometrics, it is often to investigate the half-life of the citations; therefore, it is important to be able to set range within some recent years as well. (Note: VxInsight provides the flexibility of showing this function; however, we had not been able to demonstrate this to users due to the lack of knowing this function then).
- None of them had tried the query function.
- Only one subject noticed the color changes on spots showed in Galaxy landscape and asked about what the color represents.
- One subject expressed her opinions about keeping the same color code in Galaxy landscape when change the landscape from the Terrain, i.e. Red color on the top, then yellow, green, and blue, etc.

7. Conclusions

(1) Data analysis:

- To clean up the dataset needs quite a lot effort, we had encountered several problems. We had overcomed the parsing XML, generating tabbed tables, and eliminating doubled-length ids. However, the variations of name form, the inconsistent usages of punctuations are the most challenging issue, which in the real world often times people use different citation styles. Within this time frame we cannot successfully parse out the cited authors from the cited references, For future work, we can use the author(s) to cited author(s) connection to explore their relationships.
- We learned the similarity measures through the tests of Bibexcel. Its step-by-step functions, tedious yet helps us to understand the concept of similarity measures of bibliographic coupling and co-citation. The Bibexcel experience with the ease of use of VxInsight together build up our knowledge in data mapping.

- Users need to be familiar with the content of the dataset so that they will be more interested to learn more about the tool and can easily comprehend the map. We had done this comparison by displaying the InfoVis dataset and the citations dataset of Marcia Bates' Berrypicking theory to library, archival and information science faculty and students. Our results showed that even though the display is the same; however, the Bates' dataset generates more questions from those users.

- InfoVis data inspired us to collect the course materials dataset to see how InfoVis courses in different universities are using these papers to support their teaching. Do they tend to use the similar materials or different? We collected course materials from InfoVis courses from three universities: UBC, Berkeley, and Stanford. The result shows that there is some overlap, but not a very significant one (see Figure 9).

(2) Software tools:

VxInsight is a commercial tool, its capability has been gone through many tests by scientists and researchers. It is very easy to install with a well written user guide, and several published papers to allow us to absorb the idea and deal with the questions about its display and usage.

We like the Pathfinder Network (White, 2003) very much for its ability to retain the most meaningful edges.

(3) Through out the process of exploring the dataset and evaluating VxInsight, we found ourselves working backwards. We often learn about patterns and phenomenon in the dataset by not using VxInsight, but other tools like MS Excel. For example, we use Excel to manipulate the dataset and produce a ranking of the top 10 most cited authors, then we try to verify this pattern using VxInsight. In this case, this means including a new relational database table "cited reference" to allow researchers who are cited by a particular cluster of papers to be displayed as the labeling for those papers. This way, we can see clearly who are cited, and can even use the slider to further explore the evolution of cited authors through time.

Perhaps this means that VxInsight, while being easy to learn, is hard to master. The initial 3D view that we obtain from VxInsight is not high in information density – we

see clusters of papers being displayed but we couldn't tell much from it, since the peak labeling that was available to us merely shows the author of the papers under the mountain peak. It actually takes some fiddling with the dataset outside of VxInsight to make us realize what we want to show as peak labels. One great lesson here is to be careful with selecting peak labeling, as the labels often end up providing the most insight, when compared to the flashy 3D view. Particularly, it is a good idea to include cited authors as a labeling in our case.

(4) Cooperation between two different mind sets: Ron is a master student in computer science and Chia-Ning is a doctoral student in library & information science, we both come from Taiwan; therefore, we have no problems in communication; however, sometimes there were gaps between us due to different interpretations of technical terms from our different background. It is a great learning experience to accomplish this assignment together.

(5) Finally, we want to thank Mr. Brian Wylie of VisWave Company, who is also one of the authors and creators of VxInsight, that he has been very supportive to help us solving out the problems we had.

8.   Appendices

Appendix A:

There are all together 17 different files we had generated successfully as shown in the following table (Table 1):

1	file.doc	The original bibliographic records as a text file with .doc attribute (not Microsoft's Word file). It requires TAB delimiters and "|" (spike) at the end of each field, and "||" at the end of a record.
2	file.out	Fields selected from file.doc and saved as a new file with .out attribute.
3	file.cit	Read file.out and sort by frequency
4	file.err	File.doc after running file check found errors
5	file.oux	A file with frequency calculated, sort, remove duplicates, and make new out file.
6	file.nmr	a file with article ids.
7	file.num	Take a list of article_ids from file.nmr and save into a new file
8	file.coc	a file contains co-occurrence frequency + cluster pairs
9	file.per	a file contains cluster pairs
10	file.add	add author names in citing-cited cluster pairs
11	file.unt	a file contains the units (e.g., authors)
12	file.uot	a shorter file made from .unt
13	file.cou	bibliographic coupling results (two articles cited very similar references)
14	file.lin	This file has the article_id in the first column and the ref_id in the second column
15	file.mul	a file with "units per record" analyzed
16	file.mut	a file with no. of unites of each article analyzed
17	file.ma2	a co-occurrences file in matrix form

Table 2: Bibexcel – File types

Appendix B. **Top ten the most cited references.**

Rec_id	No. cited	Cited author(s)	Cited Title	Source	Year
acm108883	65	Stuart K. Card	Cone Trees: animated 3D visualizations of hierarchical information	Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, April 27-May 02, New Orleans, Louisiana, United Statesp.189-194	1991
acm22342	61	G. W. Furnas	Generalized fisheye views	Proceedings of the SIGCHI conference on Human factors in computing systems, April 13-17, Boston, Massachusetts, United States, p. 16-23	1986
acm108870	37	,J. D. Mackinlay, G. G.Robertson, Stuart K.Card,	The Perspective Wall Detail and context smoothly integrated	Proceedings of CHI '91	1991
acm33404	35	Edward R. Tufte	The visual display of quantitative information	Graphics Press, Cheshire, CT.	1986
acm223956	28	John Lamping Ramana Rao Peter Pirolli	A focus+context technique based on hyperbolic geometry for visualizing large hierarchies	Proceedings of the SIGCHI conference on Human factors in computing systems, May 07-11, Denver, Colorado, p. 401-408.	1995
acm153577	25	George G. Robertson Stuart K. Card Jack D. Mackinlay	Information visualization using 3D interactive animation	Communications of the ACM, v.36 n.4, April 1993, p.57-71.	1993
acm191775	24	Christopher Ahlberg Ben Shneiderman	Visual information seeking: tight coupling of dynamic query filters with starfield displays	Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, Boston, Mass., April 24-28, p. 313-317.	1994
acm949654	24	Brian Johnson Ben Shneiderman	Tree-Maps: a space-filling approach to the visualization of hierarchical information structures	Proceedings of the 2nd conference on Visualization '91, October 22-25, San Diego, California. (IEEE)	1991
acm22950	23	Jock Mackinlay	Automating the design of graphical presentations of relational information	ACM Transactions on Graphics (TOG), v.5 n. 2, April 1986, p. 110-141.	1986
acm108874	22	Stuart K. Card George G. Robertson Jock D. Mackinlay	The information visualizer, an information workspace	Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, April 27-may 02, New Orleans, Louisiana, p. 181-186.	1991

8. Bibliography

Benford, Steve (1995) "Information Visualisation, Browsing and Sharing in Populated Information Terrains," In *Proceedings of the Seminar Series on New Directions in Software Development : The World Wide Web*. University of Wolverhampton, March 1995.

Boerner, K. (2000). "Extracting and visualizing semantic structures in retrieval results for browsing," ACM Digital Libraries ' 00, San Antonio, TX, June 2000.

Boerner, Katy, Chen, Chaomei, and Boyack, Kevin W. (2003) "Visualizing Knowledge Domains," Annual Review of Information Science and Technology," 37, 2003, p. 178-255.

Boyack, K.W., Wylie, B.N., Davidson, G.S. & Johnson, D.K. (2000). "Analysis of patent databases using VxInsight," ACM New Paradigms in Information Visualization and Manipulation, McLean, VA, Nov. 10, 2000.

Boyack, Kevin W., Wylie, Brian N., and George S. Davidson (2001a) "A Call to Researchers: Digital Libraries Need Collaboration Across Disciplines," D-Lib Magazine, 7(3), Oct. 2001. http://www.dlib.org/dlib/october01/boyack/10boyack.html

Boyack, Kevin, W., Wylie, Brian N., and Davison, George S. (2001b) "Information Visualization, Human-Computer Interaction, and Cognitive Psychology: Domain Visualizations," Joint Conference on Digital Libraries, June 2001.

Boyack, Keven W., Wylie, Brian N., and Davidson, George, S. (2002) "Domain Visualization Using VxInsight for Science and Technology Management," Journal of the American Society for Information Science and Technology, 53(9), 2002, p. 764-774.

Boyack, Kevin W. and Boerner, Katy (2003) "Indicator-Assisted Evaluation and Funding of Research: Visualizing the Influence of Grants on the Number and Citation Counts of Research Papers," Journal of the American Society for Information Science and Technology, 54(5), 2003, p. 447-461.

Chen, C. (1999) "Visualising semantic spaces and author co-citation networks in digital libraries," Information Processing and Management, 35, 1999, p. 401-420.

Chen, C., Paul, R.J., and O' Keefe, B. (2001a) "Fitting the Jigsaw of Citation: Information Visualization in Domain Analysis," Journal of the American Society for Information Science and Technology, 52(4), 2001, p. 315-330.

Chen, Chaomei and Paul, Ray, J. (2001b) "Visualizing A Knowledge Domain' s Intellectual Structure," IEEE Computer, March 2001, p. 65-71.

Chen, Chaomei and Morris, Steven (2003)"Visualizing Evolving Neworks: Minimum Spanning Trees versus Pathfinder Networks," IEEE Symposium on Information Visualization 2003, October 2003.

Davidson, George S., Hendrickson, Bruce, Johnson, David K., Meyers, Charles E., and Wylie, Brian N. (1998) "Knowledge Mining with VXInsight: Discovery Through Interaction," Journal of Intelligent System, 11(3), Nov/Dec 1998, p.259-285.

Davidson, George S., Wylie, Brian N., and Boyack, Kevin W. "Cluster Stability and the Use of Noise in Interpretation of Clustering," Proceedings IEEE Information Visualization 2001, p. 23-30.

Hearst, Marti and Karadi, Chandu (1997a) "Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy," *in Proceedings of the 20th Annual International ACM/SIGIR Conference,* Philadelphia, PA, July 1997.

Hearst, Marti and Karadi, Chandu (1997b) "Searching and Browsing Text Collections with Large Category Hierarchies," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Atlanta, GA, March 1997.

Hetzler, B., Whitney, P., Martucci, L., & Thomas, J. (1998). "Multi-faceted insight through interoperable visual information analysis paradigms," Proceedings of IEEE Information Visualization ' 98, p. 137-144.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998) "Introduction to Latent Semantic Analysis," Discourse Processes, 25, 259-284.

Nederhof, A.J., and E. Van Wijk (1997) "Mapping the social and behavioral sciences world-wide: Use of maps in portfolio analysis of national research efforts," Scientometrics, 40, 1997, p. 237-276.

Newman, M. E. J. (2001) "Scientific Collaboration Networks, I: Network Construction and Fundamental Results," Physical Review E, 64, paper number 016131.

Noyons, E. C. M. & van Raan, A. F. J. (1998) "Advanced Mapping of Science and Technology," Scientometrics, 41, 1998, p. 61-67.

Small, H. (1999). "Visualizing Science by Citation Mapping," Journal of the American Society for Information Science, 50(9), July 1999, p. 799-813.

Spasser, M.A. (1997). "Mapping the Terrain of Pharmacy: Coclassification Analysis of the International Pharmaceutical Abstracts database," Scientometrics, 39, p. 77-97.

Werner-Washburne, Margaret, Wylie, Brian, Boyack, Kevin, Fuge, Edwina, Galbraith, Judith, Weber, Jose, and Davidson, George (2002) "Comparative Analysis of Multiple Genome-Scale Data Sets," Genome Research, 12(10), 2002 Oct; 1564-73.

White, Howard D. (2003) "Pathfinder Networks and Author Cocitation Analysis: A Remapping of Paradigmatic Information Scientists," Journal of the American Society for Information Science and Technology, 54(5), 2003, p. 423-434,

White, H. D. and McCain, K. W. (1998) "Visualizing a Discipline: An Author Co-citation of Information Science, 1972-1995," Journal of the American Society for Information Science and Technology, 53(9), 2002, p. 764-774.

Wise, J.A. (1999). "The Ecological Approach to Text Visualization," Journal of the American Society for Information Science 50(13), 1224-1233.