Major Information Visualization Authors, Papers and Topics in the ACM Library

Weimao Ke, Katy Börner and Lalitha Viswanath
Indiana University, School of Library and Information Science & School of Informatics
10th Street & Jordan Avenue, Main Library
Bloomington, IN 47405, USA
{wke;katy;lviswana}@indiana.edu

Abstract

The presented work aims to identify major research topics, relationships between members of the community, and trends over time in the IV Contest 2004 data set. Co-author, paper-citation, word co-occurrence and burst analyses were used to analyze the data set. The results are visually presented as sample graphs, static Pajek [1] visualizations and interactive network layouts using prefuse [2]. The complementary web page with details on and raw data of all underlined data sets is available at http://ella.slis.indiana.edu/~lviswana/InfoVis/iv2004.htm.

The InfoVis Contest Data Set

The InfoVis Contest dataset contains 614 papers that were published between 1974 and 2004. The papers come with a title, authors, abstracts, keywords, source, references, number of pages, and year of publication. One paper (acm673478) has no author. 429 papers have abstracts, 424 papers come with keywords and 340 papers have an abstract and keywords. The yearly distribution is plotted in Figure 1.
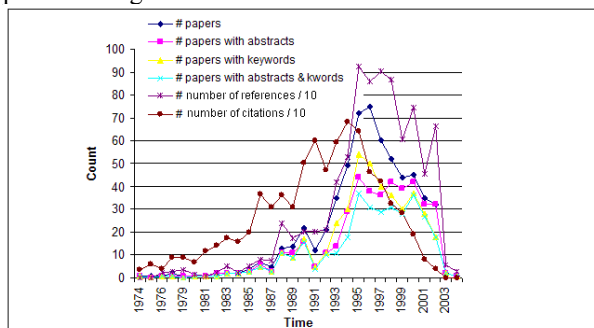
Figure 1. Yearly distribution of papers, abstracts, keywords, references and citation counts.

The data set has a total number of 8502 unique references, with 5 duplicates. Out of those, 1970 link to papers within the contest data set, called IV core. 1801 references link to other ACM papers and 4722 link to non-ACM papers.

We identified 106 unique publication venues. Unfortunately, these venues do not include the annual Information Visualization Conference in London, the annual SPIE Visualization and Data Analysis Conference in San Jose, or the new *Information Visualization* journal. Hence all subsequent analysis will provide a partial view of InfoVis research and education.

In preparation for the subsequent analysis, the total set of 1161 authors was reduced to 1036 unique authors and duplicate author ids were merged. 1859 keywords were reduced to 1753 unique keywords. Publication years were successfully retrieved for 8178 out of the 8502 references.

TASK 1: Static Overview of 10 Years of InfoVis

Knowledge domain visualization techniques [3] were applied to map the semantic space of the data set via citation analysis and co-citation analysis. The results of the citation analysis are visualized in Pajek [1] and are shown in Figure 2.
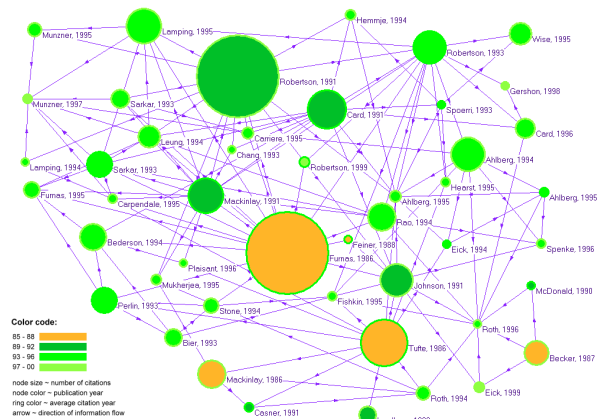
Figure 2. Citation network

Depicted are papers that got cited at least 20 times (15 papers) and all the papers that cited those papers and themselves got cited at least 7 times (44 papers). Elimination of duplicate entries resulted in 47 papers. Each paper is represented by a circle. Node size denotes the number of received citations. Node color denotes year of publication. Ring color denotes the average citation year. Links represent direct citation links.

Within IV core there are two papers that received 70 citations: Furnas's 1986 paper entitled *Generalized fisheye views*, and Robertson's 1991 paper *Cone trees: animated 3D visualizations of hierarchical information*. Tufte's 1986 paper *The visual display of quantitative information* was cited 40 times (see article_cited_count_withinset).

It is interesting to note that Bertin's 1983 paper on the *Semiology of Graphics* is cited most often (14 times) among papers in the non-ACM category. It is followed by Spence & Apperley's 1982 paper *Database Navigation*

An Office Environment for the Professional, which has 9 citations (see article_cited_count_outside_acm).

TASK 2: Major Research Topics and Their Evolution

Sudden increases in the usage frequency of keywords were identified using Kleinberg's burst analysis algorithm [4]. The results for unique keywords (several are compound terms) are shown in Figure 3.

Word	Burst_Weight	Burst_Years
Data visualization	3.7	(1994 - 1995)
focus+context	4.29	(1999 - 2002)
hierarchy	3.95	(2000 - 2002)
human factors	3.42	(1983 - 1994)
information visualization	13.083	(1998 - present)
user interface	3.457	(1983 - 1991)

Figure 3. Keyword burst analysis results

While InfoVis research seems to have started with *user_interface* and *human_factors* research, *data_visualization* work dominated in 94/95. *Information Visualization* has a strong, still ongoing burst since 1998.

TASK 3: The Authors in the InfoVis Contest Set

Scholars with more than 10 papers are Ben Shneiderman(23), Stuart K. Card (16), Jock D. Mackinlay (15), Steven F. Roth (12), George Robertson(11), Daniel A. Keim (11) and John T. Stasko (11).

Authors that received more than 100 citation links are Stuart K. Card (236), Jock D. Mackinlay (212), George G. Robertson (180) and Ben Shneiderman (173).

The top four authors with the largest number of unique co-authors are Ben Shneiderman(23), Stuart K. Card (17), Jock D. Mackinlay(17) and George G. Robertson (16).

In IV core, 93.3% of the authors have co-authored. Figure 4 shows a co-author network for IV core authors that published no less than 10 papers OR got cited no less than 50 times OR have no less than 20 times of co-authorship with other authors. 17 authors satisfied one or more of the three criteria. All of their co-authors are shown as well resulting in 138 author nodes. The node size corresponds to the number of papers published. Node color denotes the total number of received citations. Edge thickness indicates the number of times authors co-authored together. The visualization reveals that Ben Shneiderman has authored the most papers (23) while Stuart K. Card received the most citations for his work. Diverse clusters of co-authors can be identified and are discussed in the accompanying web page.
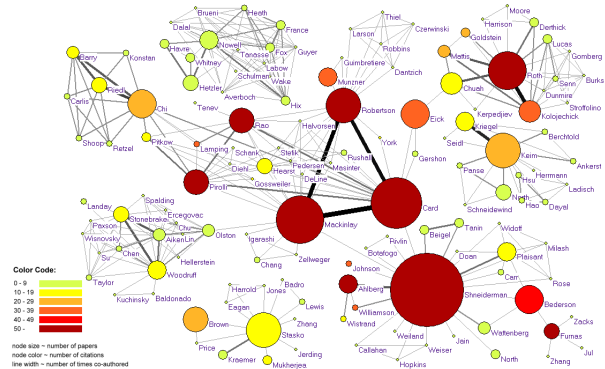

Figure 4. Co-author space of highly productive, cited or co-authoring authors.

Discussion

We presented simple statistics, burst analysis results of keywords and semantic maps of major papers and authors based on the InfoVis Contest 2004 data set.

Given that the data set does not cover papers presented at the annual InfoVis Conference in London or the annual SPIE Visualization and Data Analysis Conference in San Jose, or the new *Information Visualization* journal, or books, only a partial picture of the domain could be drawn.

Acknowledgements

References

1. Batagelj, V., & Mrvar, A. *Pajek: Program Package for Large Network Analysis, University of Ljubljana, Slovenia*. 1997.
2. Kleinberg, J.M. (2002). *Bursty and hierarchical structure in streams*. In *8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*. ACM Press.
3. Börner, K., Chen, C. & Boyack, K. (2003).*Visualizing Knowledge Domains*. In *Annual Review of Information Science & Technology*, B. Cronin (Ed.) Information Today, Inc./American Society for Information Science and Technology: Medford, NJ. p. 179-255.
4. Börner, K., Maru, J. & Goldstone, R. (2004). The simultaneous evolution of author and paper networks. PNAS, 101(Suppl_1):5266-5273.
5. Heer, J., Card, S. K., & Landay, J. A. (2004). *Prefuse: A toolkit for interactive information visualization.* Submitted to *User Interface and Software Technology*, April 2004.