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

Weimao Ke[*]
School of Library and Information Science
Indiana University

Katy Börner[+]
School of Library and Information Science
Indiana University

Lalitha Viswanath[‡]
School of Informatics
Indiana University

ABSTRACT

The presented work aims to identify major research topics, co-authorships, and trends in the IV Contest 2004 dataset. Co-author, paper-citation, and burst analysis were used to analyze the dataset. The results are visually presented as graphs, static Pajek [1] visualizations and interactive network layouts using Pajek's SVG output feature. A complementary web page with all the raw data, details of the analyses, and high resolution images of all figures is available online at http://iv.slis.indiana.edu/ref/iv04contest/.

Keywords: Citation analysis, co-author analysis, visualization

1   THE INFOVIS CONTEST DATASET

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.
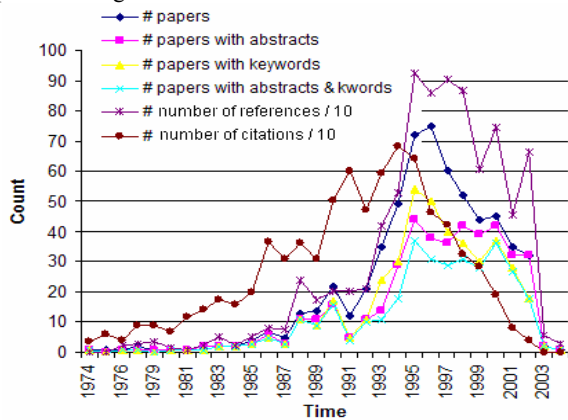

Figure 1. Yearly distribution of papers, abstracts, keywords, references and citation counts.

The dataset has a total number of 8,502 unique references. Out of these, 1,970 link to papers within the contest dataset, called the IV core. 1,801 references link to other ACM papers and 4,722 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, the subsequent analysis will provide only a partial view of InfoVis research and education.

During data cleaning, the total set of 1,161 authors was reduced to 1,036 unique authors and 1,859 keywords were reduced to 1,753 unique keywords. The year of publication was successfully retrieved for 8,178 out of the 8,502 citation references.

2   TASK 1: STATIC OVERVIEW OF 10 YEARS OF INFOVIS

Knowledge domain visualization techniques [2] were applied to map the semantic space of the dataset via citation analysis. The results of the citation analysis were visualized with Pajek [1] and are shown in Figure 2.
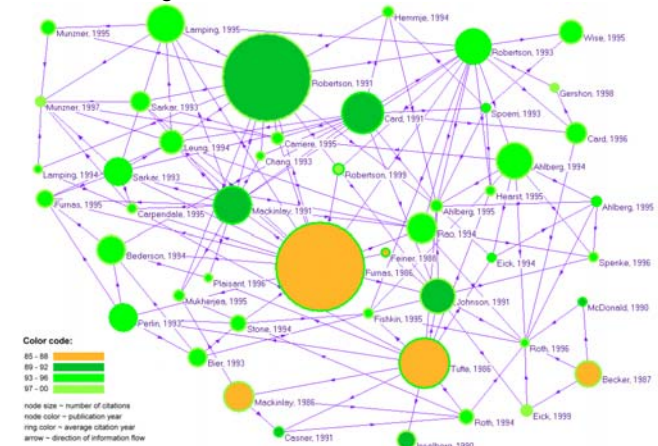

Figure 2. Citation network of major papers

Depicted are papers that got cited at least 20 times (15 papers) as well as papers which were cited at least 7 times AND that cited those 15 papers (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 the IV core there are two papers that received 70 citations: Furnas's 1986 paper entitled *Generalized fisheye views*, and Robertson's 1991 paper on *Cone trees: animated 3D visualizations of hierarchical information*. Tufte's 1986 paper entitled *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).

---

* email: wke@indiana.edu
+ email: katy@indiana.edu
‡ email: lviswana@indiana.edu

3 TASK 2: MAJOR RESEARCH TOPICS AND THEIR EVOLUTION

Sudden increases in the usage frequency of keywords were identified using Kleinberg's burst analysis algorithm [3]. The results for <u>unique keywords</u> (several are compound terms) are shown in Table 1.

Table 1. Keyword burst analysis results

Word	Burst_Weight	Burst_Years
Data visualization	3.70	(1994 - 1995)
focus+context	4.29	(1999 - 2002)
hierarchy	3.95	(2000 - 2002)
human factors	3.42	(1983 - 1994)
information visualization	13.08	(1998 - present)
user interface	3.46	(1983 - 1991)

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

4 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 <u>number of unique co-authors</u> are Ben Shneiderman (23), Stuart K. Card (17), Jock D. Mackinlay (17) and George G. Robertson (16).

In the IV core, 93.3% of the authors have co-authored. Figure 3 shows a co-author network for the IV core authors that published no less than 10 papers OR got cited no less than 50 times OR have no less than 20 occurrences of co-authorship with other authors. 17 authors satisfied one or more of the three criteria. All of their co-authors are shown, resulting in 138 author nodes.
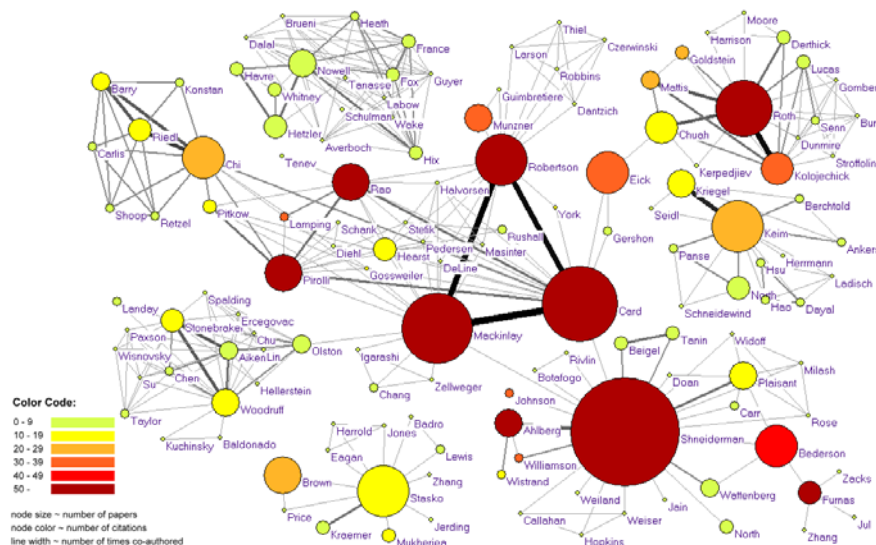
The node size corresponds to the number of papers published. Node color denotes the total number of received citations. Edge thickness indicates the number of times authors co-authored together. The visualization reveals that Ben Shneiderman has authored the most papers (23), while Stuart K. Card received the most citations for his work. Diverse clusters of co-authors can also be identified and are discussed in the accompanying web page.

5 DISCUSSION

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

Given that the dataset does not cover papers presented at two of the three main annual InfoVis Conferences, the new *Information Visualization* journal, or books, only a partial picture of the domain has been drawn.

REFERENCES

1. Batagelj, V. and A. Mrvar, *Pajek: Program Package for Large Network Analysis, University of Ljubljana, Slovenia*. 1997, http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
2. Börner, K., C. Chen, and K. Boyack, *Visualizing Knowledge Domains*, In *Annual Review of Information Science & Technology*, B. Cronin, Editor. 2003, Information Today, Inc./American Society for Information Science and Technology: Medford, NJ. p. 179-255.
3. Kleinberg, J.M. *Bursty and hierarchical structure in streams*. In *8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*. 2002: ACM Press. p. 91-101.

Figure 3. Co-author space of highly productive, cited or co-authoring authors.