

Understanding Eight Years of InfoVis Conferences using PaperLens

Bongshin Lee*
University of Maryland
Microsoft Research

Mary Czerwinski†
Microsoft Research

George Robertson‡
Microsoft Research

Benjamin B. Bederson§
University of Maryland

ABSTRACT

We present PaperLens, a visualization that reveals connections, trends, and activity throughout the InfoVis conference community for the last 8 years. It tightly couples views across papers, authors, and references. This paper describes how we analyzed the data, the strengths and weaknesses of PaperLens, and interesting patterns and relationships we have discovered using PaperLens.

1 INTRODUCTION

Information Visualization has been studied extensively over the last 10 years to help people explore and understand data. From 1995 to 2002, 315 authors published 155 papers in the InfoVis symposium. Interfaces for visualizing search results for a digital library, such as Envision [2], exist, but we do not yet have a visualization system that allows researchers in our field to understand how researchers, topics, and outside research sources interact and influence research activity in general. In order to address this issue, we have developed a visualization called PaperLens that supports the discovery and identification of major research topics, relationships between members of the community, and trends over time (Figure 1).

2 DATA ANALYSIS

2.1 Topic Clustering

We used technology developed internally at Microsoft Research to cluster the papers in the InfoVis proceedings. The clustering software was originally developed for site administrators to help build and maintain category hierarchies for documents. The text-clustering component suggests a hierarchically organized set of categories when no such structure exists. In order to cluster, we used the titles, references and keywords (if available) in the clustering process, weighting the titles more heavily to get a better clustering result. A standard list of stop words, months of the year, journal and proceeding titles and version and page numbers were removed from influencing the cluster results. Five clusters emerged from using this tool:

- General (58 papers)
- Dynamic Queries (28)
- Graph Visualization (19)
- Focus + Context Techniques (16)
- Tree Visualization (31)

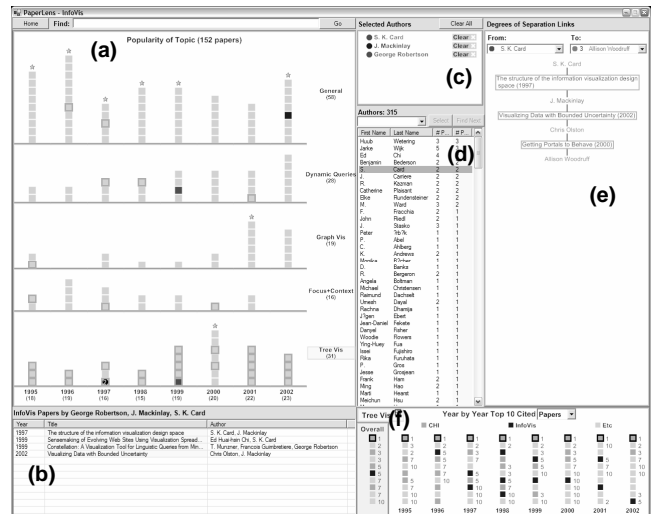


Figure 1. PaperLens: (a) Popularity of Topic (b) Paper List (c) Selected Authors (d) Author List (e) Degrees of Separation Links (f) Year by Year Top 10 Cited Papers/Authors.

2.2 Co-authorship Analysis

A co-author collaboration graph is often used to find the relationship between authors and the center of the community, i.e., the author that has the shortest average path length to all other authors in the graph [3]. However, the graph among InfoVis authors is too fragmented to give any useful insights. S. F. Roth, the center of the graph, has published 5 papers with 13 co-authors and has only 19 related colleagues among 315 possible individuals. We instead decided to display all of the related colleagues when an author is selected by the user. We compute the shortest path length between two authors on demand and call it *degrees of separation*.

2.3 Reference Citation Counts

One of the interesting questions we wanted to answer was “Which papers/authors are most often referenced?” because this is one important metric indicating influential papers/authors. In addition to counting the number of references overall, we computed them by year and by topic to show their trends.

3 STRENGTHS

3.1 Evolution of Topics

We can easily capture trends of the topics because we organized papers by their topic and signified the most popular topic each year by a small star above the relevant column in the *popularity of topic* view (Figure 1a). For example, the topic of Graph Visualization has grown in popularity quite recently and was most popular in 2001, while the topic of Dynamic Queries has exhibited a steady increase over the last 8 years.

*email: bongshin@cs.umd.edu

† email: marycz@microsoft.com

‡ email: ggr@microsoft.com

§ email: bederson@cs.umd.edu

3.2 Easy Access to Papers/Authors

PaperLens enables users to get a list of papers by year, by topic, or by authors. It also provides a way to search for specific papers/authors. By clicking on a year, a fisheye technique helps to reveal the individual paper titles for that year by topic. By selecting a topic in the display, the list of all papers in that area is shown in the *paper list* (Figure 1b). When the user selects authors from the authors list, they are shown in the *selected authors* area (Figure 1c) and added to the *From* combo-box in the *degrees of separation links* view (Figure 1e). Once authors are added to the *selected authors* area, all of the papers by them are also shown in the *paper list* and are highlighted in the *popularity of topic* view, matched to the author by color coding. When two or more selected authors wrote a paper together, we used black and showed the number of authors from the selected authors on the paper. The user can iterate through each author's papers by clicking on the author name or by clicking on them in the *paper list*.

When the user selects a paper by single click, its authors are automatically picked from the *author list* (Figure 1d) and added to the *selected authors* area. A double-click takes the user to the ACM Portal with a link to the paper.

3.3 Most Frequently Published Authors

We show the number of papers published by an author in the *author list* (Figure 1d). Users can sort the *author list* by # of papers and immediately see who has published most. By selecting a topic in the *popularity of topic* view, a column having the number of papers in that topic is added to the *author list*. Now users can see who has published most in each topic.

3.4 Relationships between Authors

The *degrees of separation links* view plays an important role in showing the relationships between authors in the InfoVis community. Once an author is selected from the *From* combo-box, we display all of the related colleagues to the selected authors in the *To* combo-box (Figure 1e) with the corresponding degrees of separation. Once an author is selected from the *To* combo-box, we display the shortest path between two authors in our dataset through their *degrees of separation links*. For example, S. K. Card and Allison Woodruff are connected indirectly to each other by 3 papers and two co-authors, Chris Olston and J. Mackinlay.

3.5 Most Frequently Referenced Papers/Authors

The top 10 overall most frequently cited references, which may or may not be from the InfoVis proceedings, are viewable in the *Overall* list (Figure 1f). One can also see the detailed information for each of the top 10 papers and how it was referenced over time by hovering on an individual paper (Figure 2a). It is easy to see that the most frequently cited paper by InfoVis authors to date is the Cone Trees paper by Robertson *et al.*, which was published in the CHI proceedings (Figure 2a). In fact, we distinguished the most influential source of cited papers (by number of citations) with the color green and the important InfoVis papers with blue. It can immediately be seen that CHI publications have been the main source of inspiration for the InfoVis community to date and that, in the top 10 papers most frequently cited papers, there is only one InfoVis publication.

PaperLens also shows frequently cited authors that either have or have not published in InfoVis. Selecting an author from the *year by year top 10 cited authors* view (Figure 1f) shows not only any papers selected authors have published in InfoVis, but also those papers that have referenced them via orange highlighting in

the *popularity of topic* area. The user can immediately discover which areas were most influenced by the selected author.

When the user selects a topic from the *popularity of topic* view, the *year by year top 10 citations* area is filtered to only show the frequent citations for that topic area. For instance, for Graph Visualization, Peter Eades, who first introduced the spring model to graph drawing, was the #1 most frequently cited author (Figure 2b), but he has not published in InfoVis as of yet. In this way, PaperLens allows the user to quickly discover the most influential researchers in a particular topic area, in addition to gaining easy access to their papers and colleagues.

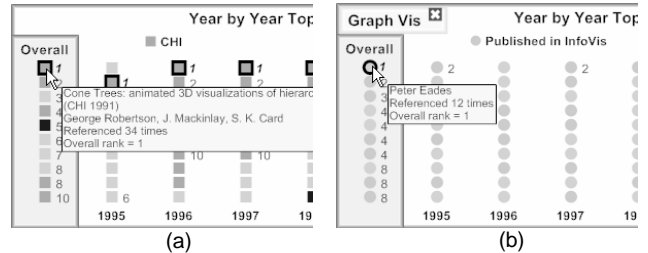


Figure 2. (a) Cone Trees paper is the most referenced paper by all InfoVis papers and (b) Peter Eades is the most frequently cited author for the Graph Visualization topic.

4 WEAKNESSES

As is often the case with powerful and new visualization tools, PaperLens requires some learning time. The users need to learn how to decode the various color mappings and highlighting. They are also required to understand how views are coupled because all views can be used for both input and output.

5 CONCLUSIONS

Using PaperLens, we have discovered many interesting patterns and relationships, which could not have been revealed using existing tools. For example, CHI was the most influential source of references used by InfoVis authors to date. George Robertson, and George Furnas, both influential in the InfoVis proceedings, have only published once or never, respectively, at InfoVis. PaperLens' power comes from its tightly coupled views across papers, authors, and references. The ability to query by time and topic has enabled novel views that we have found to be very beneficial in our explorations of this domain. As future work we are examining ways to scale the visualization to a larger dataset of documents with many other kinds of metadata.

ACKNOWLEDGEMENTS

PaperLens was built using Piccolo.NET, a port of Piccolo.Java [1]. We would like to thank Aaron Clamage who ported .NET versions of Piccolo.

REFERENCES

- [1] Benjamin B. Bederson, Jesse Grosjean, and Jon Meyer. "Toolkit Design for Interactive Structured Graphics", *IEEE Transactions on Software Engineering*, 2004, pp. 535-546
- [2] Lucy Terry Nowell, Robert K. France, and Deborah Hix. "Exploring Search Results with Envision", *Proceedings of CHI, 1997*, pp. 14-15.
- [3] Alan F. Smeaton, Gary Keogh, Cathal Gurrin, Kieran McDonald, Tom Sødring, "Analysis of papers from twenty-five years of SIGIR conferences: what have we been doing for the last quarter of a century?", *ACM SIGIR Forum*, 2003, pp. 49-53.