CiteVis: Exploring Conference Paper Citation Data Visually

John Stasko*, Jaegul Choo, Yi Han, Mengdie Hu, Hannah Pileggi, Ramik Sadana, Charles D. Stolper
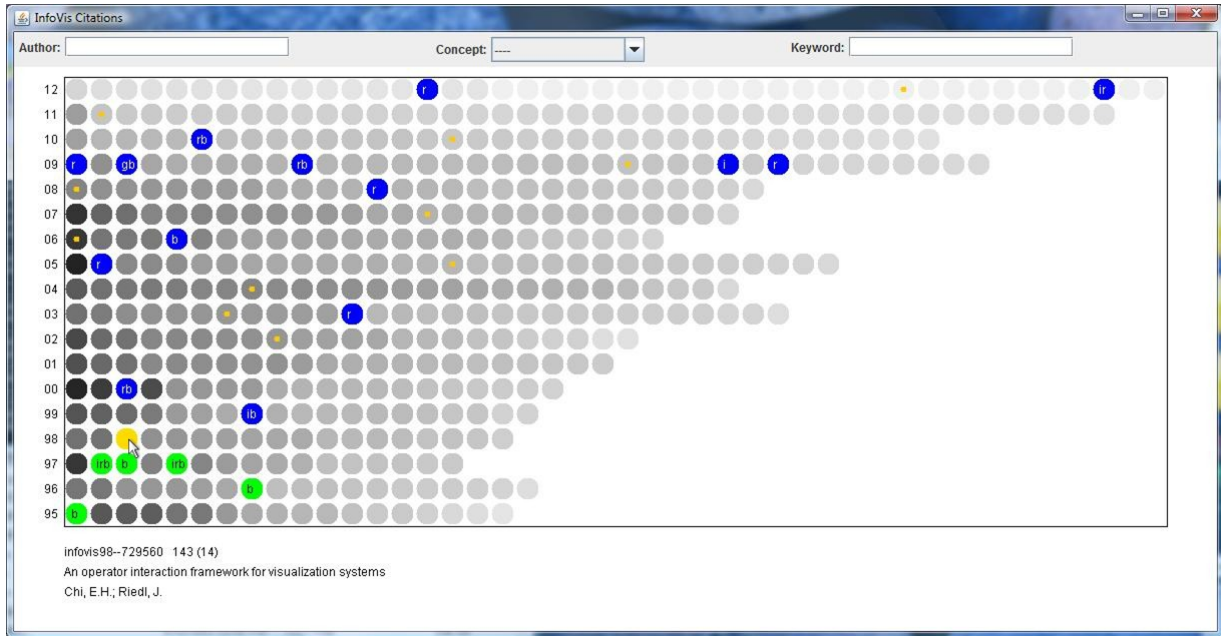
Georgia Institute of Technology

Figure 1. The user interface for CiteVis with a focus paper (yellow), papers citing it (blue), and its citations (green). The system can be found at http://www.cc.gatech.edu/gvu/ii/citevis.

ABSTRACT

Citation counts and intra-conference citations are one useful measure of the impact of prior research in a field. We have developed CiteVis, a visualization system for portraying citation data about the IEEE InfoVis Conference and its papers. Rather than use a node-link network visualization, we employ an attribute-based layout along with interaction to foster exploration and knowledge discovery.

Keywords: citation network, network visualization, interaction.

1 INTRODUCTION

The citations that an academic paper receives are one indication of the impact of that the paper in its research community. Whenever a paper cites another paper, presumably the cited paper has been influential in the subsequent research. In this work, we seek a good way to explore and understand the citation counts and patterns of all the papers that have been presented at the IEEE Information Visualization (InfoVis) Conference. We wanted to examine both the total citation count of papers and the purely internal citations within the InfoVis Conference. We also wanted to examine the citation data of particular author's papers and of papers about specific topics. To achieve these goals, we designed and built an interactive visualization system called CiteVis.

*email:{stasko | joyfull | yihan | mengdie.hu | hpileggi3 | rsadana3 | chadstolper}@gatech.edu

Previous research has sought better ways to understand conference publication and citation patterns. MacKenzie explored citation counts and patterns of HCI papers in the context of papers' and authors' impact [1]. His analyses were more statistical and summative in nature rather than focusing on individual papers. The Citeology system shows 28 years of ACM CHI and UIST conference papers, employing a node-link overview visualization with curved edges (citations) between papers [2]. The system emphasizes showing paper descendants, ancestors, and the shortest citation path between papers. While the system's visualization is very evocative and beautiful, its node-link representation often leads to the "ball of string" style view so common in many network visualizations. We believe that other representations without such a mass of edges may actually be more useful for exploration and analysis.

To gather the data for the visualization, we began with a datafile of all the InfoVis papers, including titles, abstracts, authors, and keywords. We created a list of 83 information visualization "concepts" and searched for each in a paper's title and abstract. To this data we manually added the Google Scholar citation count of each paper as of January 2013. We also manually logged each "internal" citation between two InfoVis papers and the section(s) of the paper that this citation occurred in (introduction, related work, and/or body). For these internal citations, we only included the actual conference paper itself; when a longer version of a paper was produced for a journal article, we did not include that as a target for an internal citation.

2 VISUALIZATION DESIGN

The CiteVis visualization design, depicted in Figure 1, uses an attribute-based network layout as employed by systems such as

Semantic Substrates [3]. Each small circle represents a single paper from the conference. These circles are arranged in rows that denote each year of the conference. Rather than drawing edges between the nodes to depict direct citations between papers, CiteVis uses interaction to focus on a paper and different colors to depict its citing and cited papers.

More specifically, the central rectangular region of Figure 1 holds all of the conference papers from InfoVis. Each year's papers are arranged in a horizontal row with the first year (1995) at the bottom and the most recent (2012) at the top. Notice that the more recent years have included more papers in general.

The darkness of each circle represents the total number of Google Scholar citations for the paper as of late January 2013. The individual papers within a year are arranged from the most citations (darkest) on the left to the least citations (lightest) on the right. IEEE InfoVis began identifying a "Best Paper Award" winner in 2002. CiteVis represents this award winning paper in each year since 2002 by a small yellow dot in the center of the paper's circle.

In order to learn more information about a paper and to see its citation pattern within the conference, the user can hover the mouse cursor over a paper's node. When that occurs, CiteVis displays information about the focused paper below the visualization, including its internal paper ID, its total Google Scholar citation count, its internal InfoVis citation count (in parentheses), the paper's title and list of authors. All prior papers cited by the focus paper are shown in green while all subsequent papers citing the focus paper are blue. Additionally, within the circle for each paper we depict an 'i', 'r', and/or 'b' character to denote the part of the paper that the citation occurs in: introduction, related work, and body, respectively.

In Figure 1, the user has moved the mouse over the 1998 paper by Chi and Riedl and we see the 14 papers in blue above that cite it along with the 5 papers below it in green that it cites. It has 143 total Google Scholar citations to go with the 14 internal citations.

One challenge with this design results from the reliance on interaction. When the user moves the mouse off a paper, then all the relevant citation information disappears. To address this problem, CiteVis supports the user to click the mouse on a paper which then selects that paper. All of the green (cited) and blue (citing) papers now are indicated by coloring the square region around and outside the circle of each paper. This feature allows the user to select a paper and then interactively examine its cited and citing papers to learn their identities and information.

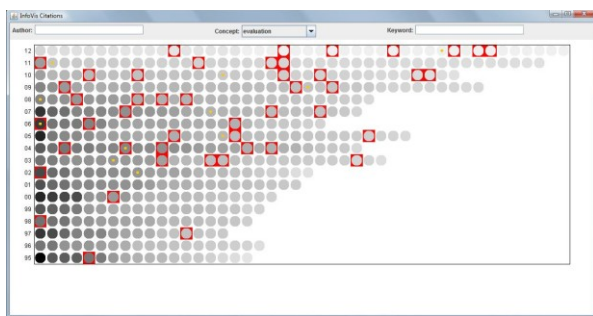CiteVis contains three input regions at the top of the window to

Figure 2. Papers mentioning "evaluation" in their title or abstract

allow the user to search for papers with specific criteria. The first and third regions allow the user to enter an author name and author-identified keyword as the focus of a search. The middle region is a pop-up menu containing 83 information visualization-related concepts that we defined, such as "interaction", "treemap",

and "network". CiteVis will search (string match) for the selected concept in the title and abstract of each paper. For each paper that matches any of these three criteria, CiteVis draws the surrounding square region of the matched papers in red. Figure 2 shows the results of searching for the concept "evaluation".

The user can press the 'i' key on the visualization to toggle the view and make the circle darkness/shading depict the number of internal InfoVis Conference paper citations. Pressing the 'e' key toggles the view back to make shading represent total external Google Scholar citations. Additionally, pressing the 'c' key clears all selections and search results.

3 DISCUSSION AND FUTURE WORK

A few sessions using CiteVis allowed us to make a number of observations about the InfoVis papers. The paper with the most external (529) and the most internal (27) citations is the "Visualizing the non-visual: spatial analysis …" paper from 1995. Other top citation papers include those about "Vizster" (2005, 397 citations), "ThemeRiver" (2000, 382 citations), and "ManyEyes" (2007, 331 citations).

The year 2000 contained four top papers with high external citation counts. Notice the four dark circles to the left edge of 2000 in Figure 1. Similarly, the year 2004 was a very strong year for papers overall. A region of darker circles extends well to the right, farther than any such similar region in other years.

Overall, papers receiving the Best Paper Award tend to be highly cited, but this varies quite a bit by year. A few are only ranked in the middle of their year.

Interacting with CiteVis we noticed a number of other insights. In an example of author focus, selecting Jeff Heer's papers showed that they tend to be highly cited, with 4 of his 12 papers at the top of their year. Certain concepts like "interaction" have been consistently strong throughout the entire period while others like "user study" and "social" have only emerged in the last 10 years. Strong technique papers such as "Hierarchical Edge Bundles" (2006) garner many citations, while design study papers in general seem to have a low number of both external and internal citations.

Like any visualization, CiteVis has a number of issues and areas for future improvement. Because the system only shows citations to the actual conference papers, some papers are underrepresented on the whole. In the years 2000-2005, before all articles went into the *IEEE TVCG* journal, authors of highly rated papers sometimes were invited to create extended versions of their papers for a journal. CiteVis does not include references to these extended journal versions in its counts.

The reliance on interaction to view citations can be both an advantage and a disadvantage. We have found that it fosters a sense of "playful" exploration, but it also makes cumulative "big picture" views more difficult.

Finally, we have numerous ideas about ways to extend the system. First, we would like to create other views such as, for example, one showing a scatterplot of normalized external versus internal citations per paper (adjusted for the age of the paper appropriately). Another view might show sub-networks of authors who cite each other consistently.

REFERENCES

[1] I. S. MacKenzie. Citedness, uncitedness, and the murky world between. *Proceedings of ACM CHI '09 Extended Abstracts*, pages 2545-2554 . ACM, May 2009.

[2] J. Matejka, T. Grossman, and G. Fitzmaurice. 2012. Citeology: visualizing paper genealogy. *Proceedings of ACM CHI '12 Extended Abstracts*, pages 181-190. ACM, May 2012.

[3] B. Shneiderman, A. Aris, Network Visualization by Semantic Substrates, *IEEE Trans. on Visualization and Computer Graphics*, 12:733-740, Sep.-Oct. 2006.