Visualization of co-authorship networks and their evolution patterns over time

MAS 965: Social Visualizations
Tripti Gore Chandorkar

This project has been an attempt to visualize co-authorship networks and their evolution patterns of three most influential authors in IEEE Information Visualization Conference over the past thirteen years (http://infovis.org/infovis2004/). The dataset of the InfoViz 2004 contest – *History of InfoViz* was used as the dataset for this project as well.

My strong interest was in the visualization of the evolution of networks that would visualize patterns and sequence of relationships and interrelationships along time. This visualization of the co-authorship networks within the InfoViz community attempts a design solution to depict time, sequence, relationships, players, movement and history in one graph. Since the theme for the project was inspired by the InfoViz 2004 contest, I looked at some of the contest winner entries to observe their approaches to the problem. I realized that none of the solutions were addressing issues of sequence and movement as I wished to do. Most of the contest visualizations were snapshots of past ten years of research in information visualization since their project scope was bigger by many folds than the scope of this project. The snapshots give a very clear overall view of the community but do not convey the working of the micro communities, events and relationships as a direct function of time as I aimed to visualize. To address this issue, this project draws from the correspondence analysis approach (Freeman, undated) with a matrix of rows depicting authors and columns depicting a range of time and papers. It produces a visual representation of the relationships between the row categories and the column categories in the same space.

Project Design

Keeping in mind my goal of visualizing patterns within co-authorship networks, I identified the most prominent authors within the community with the help of the 'article-author' and 'references' tables in the database. The number of papers (co-authored and independent) and number of citations received by an author were used as the measure for identifying the influence on other InfoViz community members and importance of the author. Accordingly, Ben Shneiderman (23 papers, 230 citations, 32 co-authors), J.Mackinlay (15 papers, 295 citations, 32 co-authors) and S. Card (16 papers, 294 citations, 32 co-authors) were identified as the top three most influential and productive authors within the InfoViz community. To begin with, these three authors and their respective co-authors were treated as three different groups. Later while mapping, patterns emerged that suggested common authors as well as exclusivity between these groups.

Most of the papers used as the source for the visualization belong to IEEE Information Visualization Conference except a few, which were authored outside the conference but are included as they have been frequently cited by other InfoViz authors.

The networks are mapped on to three different matrices. (http://web.mit.edu/tripti/www/FallMAS965/final/combination.htm). The resulting graph generates a relationship between the dimensions of co-authors, papers and time, all in the same space.

Rows:
The authors have been assigned rows according to the number of papers each one of them wrote with B. Shneiderman, J. Mackinlay and S. Card in their respective graphs.

Thickness of the horizontal gray line as well as the distance of the co-authors from the top row (the main author in the graph) indicates the strength of a tie to the main author. The strength of the tie between a co-author and the main author of a group has been assumed to be proportional to the number of papers co-authored with the main author by him/her. The size of the type indicates the prominence of the author within a group.
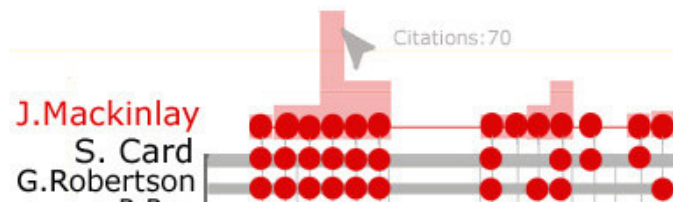
The graph also maps the level of popularity of each paper as a function of the number citations received by it.

Columns:
The columns - represent time, the last thirteen years (1992 to 2002) of the InfoViz research community and the thin vertical lines representing papers presented at the IEEE Information Visualization Conference and other important papers of the InfoViz community written by the three authors outside this conference. The number of citations received by these papers was used as a measure for calculating their importance.

For every paper in a year (a year divided into months so that more than one paper can be mapped onto one year), its corresponding co-authors (dots) are located on the vertical line representing the paper.

The dots
The dots represent the co-authors corresponding to their names on the left of the graph and their papers and the year of the paper publication. The blue dots represent B. Schneiderman and his co-authors exclusively, red dots represent J. Mackinlay and his co-authors exclusively and the green dots represent S. Card and his co-authors exclusively. The yellow dots represent co-authors within each group who have co-authored a paper independent of the main author (Shneiderman, Mackinlay or Card) while the brown dots represent co-authors common to the three groups.
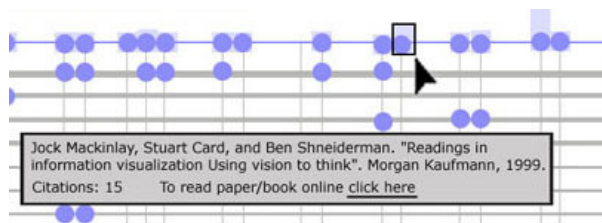

The interface also gives information about each paper if clicked on any of the dots.
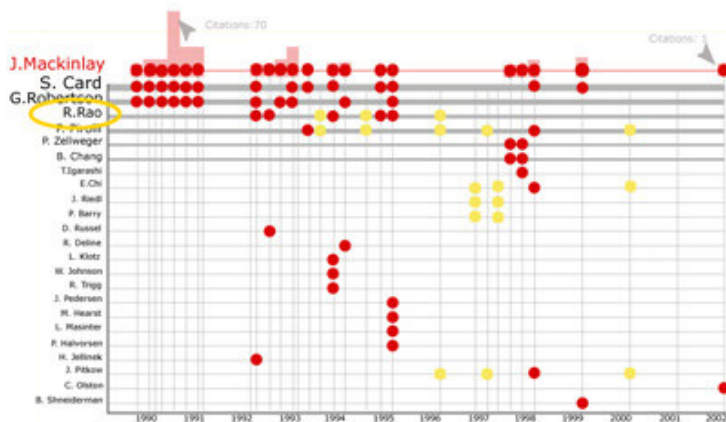
Analysis of the graphs

1. B. Shneiderman's co-authorship network works exclusively of J.Mackinlay's and S. Card's.

2. Mackinlay and Card share most of their co-authors including each other. The author closest to both of them after each other is Robertson with whom they have collaborated most often on the same papers.

3. Mackinlay, Card and Robertson are very tightly knit and work in a clique. Their collaboration has been very productive in terms of the number of papers produced and the citations received by each of them.
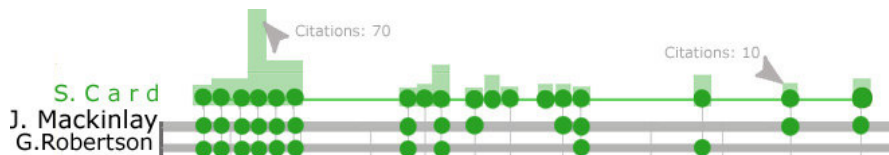
4. Shneiderman has been publishing papers very consistently over the last 13 years in InfoViz unlike Mackinlay and Card who have gaps between years of published papers. Their productivity in terms of co-authored papers has gone down due to some reason after the year 1999.

5. As evident from the thickness of the horizontal line, the tie between C. Plaisant and Shneiderman is strongest amongst all Shneiderman's co-authors in terms of number of papers produced by both of them together. The same is true for Mackinlay and Card and vice versa.

6. Although the three authors have almost the same number of co-authors, the trend shows that Shneiderman's co-authors continue to co-author with him and very few work independently of him. While in Mackinlay's and Card's groups, their co-authors show a trend to produce papers independent of him after having co-authored with him.

7. Most of Shneiderman's co-authors (in the time period of 1990 to 2002) seem to have already established ties with him to co-author papers with him directly. They do not seem to have used any stepping-stones in the form of co-authoring papers with other authors closer to Shneiderman to get closer to him.

8. In Mackinlay's and Card's graphs, Pitkow seems to have used his co-authorship with Ramana Rao and Pirolli (who are closer to Mackinlay and Card) as a stepping stone to co-author with Mackinlay and Card on a paper in 1998.
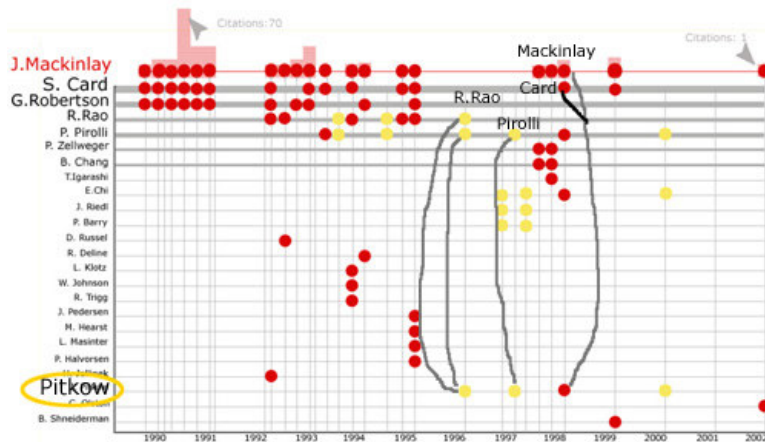
9. The only collaboration amongst the three networks is the book, " Readings in information visualization: Using vision to think", by Shneiderman, Card and Mackinlay, much later (1999) in the InfoViz time span. This point of collaboration between the three groups in the year 1999 after years of working exclusively of each other is striking. It could have many interpretations. One interpretation being that by 1999, all three had established themselves well in their respective areas and they realized each other's importance in Information Visualization a little before 1999.

10. The number of papers co-authored by Ramana Rao (the third closest co-author in the Mackinlay and Card networks) with Mackinlay and Card and without them later in the time span suggests that he is on the way to attaining one of the top author positions in the community.

11. A single glance at the graphs reveals that a paper co-authored by Mackinlay and Card in the year 1991 had the most impact in terms of citations in InfoViz out of the three authors.

Problems

To add the next layer of exploration to the analysis of the co-authorship networks, I studied the keywords of the papers of the three authors. It would be possible to read the differences or similarities in the research themes through an in-depth analysis of the keywords. It would reveal if the groups that are working independently of each other also have different research themes or if not, it would lead us to potential ties and collaboration between the three groups. Although, the keywords retrieved from the data set were different from each other and not identical per se but two keywords could be pointing to the same topic. As a newcomer to this field of research, it was difficult for me to identify these nuances hence the keyword analysis only touched the surface.

It was possible to carry out this visualization as it involved the networks of only three top authors. Much thought would be required to visualize the networks of all the authors of a research community with this correspondence mapping technique.

Conclusions and Further Work

The visualization would have been richer if the data of citations interlinking authors amongst the three groups was incorporated. This would have revealed patterns of who is citing whom and the direction of exchange of information. This part of the analysis, which would require extensive query searches in an advanced database platform, has been left out and will be carried out as a further study of the project.

Another source of data that could have been incorporated within the visualization to make it more exploratory would be the order of co-authors on the papers.

One of the goals of the visualization of these co-authorship networks is to render a form to the notion of relationships that people may have about the three authors in the InfoViz community. Being unfamiliar with the trends in the research field of information visualization, I drew the maps of networks of the three authors only by analyzing the data and without any preconceived notions of relationships. It would be interesting to get an opinion on the reality of these networks and their patterns from an insider of the InfoViz community.

References:

Freeman, L .C., (Undated), Visualizing Social Networks.

Newman, M.E.J., (Undated), Who is the best-connected scientist? A study of scientific coauthorship networks.