20 Years of four HCI conferences: A Visual Exploration

Nathalie Henry, Howard Goodell, Niklas Elmqvist, Jean-Daniel Fekete

HAL Id: hal-00851874
https://inria.hal.science/hal-00851874

Submitted on 19 Aug 2013

20 Years of Four HCI Conferences: A Visual Exploration

Nathalie Henry

INRIA/LRI, Univ. Paris-Sud & University of Sydney

Howard Goodell

INRIA/LRI, Univ. Paris-Sud

Niklas Elmqvist

INRIA/LRI, Univ. Paris-Sud

Jean-Daniel Fekete

INRIA/LRI, Univ. Paris-Sud

Abstract

We present a visual exploration of the field of human-computer interaction through the author and article metadata of four of its major conferences: the ACM conferences on Computer-Human Interaction (CHI), User Interface Software and Technology (UIST) and Advanced Visual Interfaces (AVI) and the IEEE symposium on Information Visualization (InfoVis). This article describes many global and local patterns we discovered in this dataset, together with the exploration process that produced them. Some expected patterns emerged, such as that — like most social networks — co-authorship and citation networks exhibit a power-law degree distribution, with a few widely-collaborating authors and highly-cited articles. Also, the prestigious and long-established CHI conference has the highest impact (citations by the others). Unexpected insights included that the years when a given conference was most selective are not correlated with those that produced its most highly-referenced articles, and that influential authors have distinct patterns of collaboration.

An interesting sidelight is that methods from the HCI field — exploratory data analysis by information visualization and direct-manipulation interaction — proved useful for this analysis. They allowed us to take an open-ended, exploratory approach, guided by the data itself. As we answered our original questions, new ones arose; as we confirmed patterns we expected, we discovered refinements, exceptions, and fascinating new ones.

20 Years of Four HCI Conferences: A Visual Exploration

Introduction

Peer-reviewed publications are a scientific community's fundamental mechanism of communicating and assessing its results. Therefore, studying the patterns and structure of these publications can reveal much about the community and its evolution over time. This article describes the structure of two overlapping communities: Human-Computer Interaction (HCI) and its outgrowth Information Visualization, based upon analysis of publication metadata from four of their conferences: the ACM Conference on Human Factors in Computing Systems (CHI), the ACM Symposium on User Interface Software and Technology (UIST), the ACM Working Conference on Advanced Visual Interfaces (AVI), and the IEEE Symposium on Information Visualization (InfoVis).

Performing this kind of study can benefit both members of the field itself and those who interact with them from outside. Novice researchers in HCI find a road map to its landmark research, central authors and institutions, and important trends. Experienced researchers get a global overview to help them clarify intuitions about their and their colleagues' roles in the community. Finally, to outsiders interested in evaluating researchers and programs, or scientometricians studying the methods and communities of science, such studies also provide context for comparing the HCI field to other areas of research.

Our analysis is based on data-driven visual exploration, in which the structure and content of the publication data itself has been allowed to guide the process. Whereas previous related studies usually begin with a priori questions and an expected model, we endeavor to develop our insights directly from the data. Exploratory analysis is based on several general questions: What are the global trends? What are the local trends? What

are the outliers? The great strength of exploratory analysis is its ability to raise unexpected questions. The drawback is that analysis can become a very drawn-out process, as the answer to one question raises many others that require further analysis. In this article, we describe our exploration process and provide a subset of interesting points for reflection, but we cannot hope to present a complete analysis of the field of human-computer interaction.

This article is organized as follows: We present a discussion of related work, and then describe the process of dataset collection and cleaning, our approach to visual exploration, and how the visualizations were created. The central part of the article is the actual analysis, divided into three sections: an overview of the field describing important work, key researchers and the main topics across time for the four conferences; information about how articles reference each other and the patterns of citations between authors; and the collaboration networks that compare the community structure across conferences. Finally, we discuss the lessons learned from this analysis in the context of HCI research.

Related Work

This section is a brief account of the state of the art in analyzing the publication data of scientific communities, as well as a summary of similar studies previously presented.

Publication Data and Small-World Networks

Studying the structure of a research field such as HCI is called *scientometrics*: the science of analyzing science. Scientometrics has a rich history and a dedicated journal published several times a year since 1979. The use of bibliographics or informetrics (data on publications) for scientometrics date back to 1965 (Price, 1965) and the description of informetrics in 1990 (Egghe & Rousseau, 1990). From sources such as our HCI publication dataset, several social networks can be extracted. The most studied are co-authorship

networks (networks formed by researchers authoring articles together), affiliation networks (bipartite networks of researchers and their institutions) and citation networks (networks formed by articles and their references).

Citation and co-authorship networks have been especially studied, in part because they exhibit a small-world structure (Watts & Strogatz, 1998). In (Newman, 2003), Newman presents several types of small-world networks including biological networks, social networks, information networks, and technological networks. He explains how small-world networks reflect the structure of networks in the real world.

These networks have three main properties:

- Node degree has a power-law distribution;

- The network has a high clustering coefficient, that is, it is locally dense; and

- The network has a short average distance; the average distance between any two nodes is small.

Power-law distributions are frequent in social networks. With such a distribution, the number of items with a specified rank x is $P[X = x] \approx x^{-\alpha}$ where α is a positive constant called the *exponent* of the power-law. The larger α, the more biased the distribution, with the first few items dominating the rest. In a publication network, this distribution is found in the degrees of the actors, but also in several others characteristics such as the number of citations.

The *clustering coefficient* for a vertex is the proportion of links between the vertices within its neighbor vertices divided by the number of links that could possibly exist between them.

The *short average distance* has popular applications in mathematics where the Erdös number (Goffman, 1969) is computed for each mathematician as the distance to Paul Erdös in the co-authorship network. Since 1994, the same concept has been applied for the Kevin Bacon number for actors. More recently, the Jonathan Grudin number has

been presented for the CSCW community (Horn, Finholt, Birnholtz, Motwani, & Jayaraman, 2004).

Studies and Systems

The analysis of co-authorship networks started in the mid-90s with (Kretschner, 1994; Grossman & Ion, 1995; Melin & Persson, 1996). These networks have been studied to provide information on the structure of a particular community (Newman, 2001), as well as the comparison of several communities, such as biology, physics and computer science (Newman, 2004). In the field of HCI, several studies have been published in the CSCW conferences (Horn et al., 2004; Jacovi et al., 2006) and a contest was organized for InfoVis 2004.

Most of these studies had a priori hypotheses that they evaluated by statistical methods. For example, Newman's research work mainly focuses on proving that the networks he collected are small-world networks. (Horn et al., 2004) exclusively studies the relations of CSCW researchers with the rest of the HCI fields and how they evolved with time. (Jacovi et al., 2006) is even more focused: its goal is to identify chasm articles (articles with a higher impact outside a community than within it.) None of the previous studies aimed to provide an overview of the HCI field and its important work. Also, only one was structured in a way that allowed unexpected insights: entrants in the InfoVis 2004 contest, analyzing 8 years of proceedings from the InfoVis conference (1995–2002) (Fekete, Grinstein, & Plaisant, 2004) were answering more open-ended questions and could present answers to new questions triggered by insights from the visual exploration aimed at answering the original set of questions.

For example, Ke et al. (Ke, Börner, & Viswanath, 2004) ran statistical analyzes and illustrated their findings with node-link diagrams created with JUNG to show most the important researchers and articles—filtering the dataset to obtain a readable

representation. PaperLens (Lee, Czerwinski, Robertson, & Bederson, 2004), developed by the University of Maryland and Microsoft Research, focused on interaction and simple histograms to explore statistics such as the number of articles, author centrality and topic clustering. In-Spire (Wong et al., 2004), a system created by the PNNL, produced a landscape of topics and showed their evolution. Finally, a student team from the University of Sydney worked on 3D and animated visualization of the community's evolution through time (Ahmed, Dwyer, Murray, Song, & Wu, 2004).

This article takes a broader view, analyzing and comparing the communities expressed in the data of four HCI conferences over their life spans, as well as a view of the overall community seen by combining the data. However, it uses a similar exploratory approach. We describe several stages of a breadth-first search into the data, with answers or partial answers to our first set of questions followed by another round of inquiry into the interesting questions the first exploration raised, and so on up to the limits of our available time and ingenuity.

As indicated by the information visualization contest above, visualization has recently been put to use for studying scientific communities; (Börner, Chen, & Boyack, 2003) gives an overview of relevant techniques and tools. VxInsight (Davidson, Hendrickson, Johnson, Meyers, & Wylie, 1998; Boyack, Wylie, & Davidson, 2002) is a general knowledge management system where relations between articles (i.e. citations and keywords) are used to map the data objects to a 3D terrain that is rearranged using a force-directed layout scheme. Boyack et al. used the tool to map the backbone of all scientific fields based on a large number of journal articles (Boyack, Klavans, & Börner, 2005). Similarly, CiteSpace (Chen, 2006) (recently updated to its second version) provide support for the full work process for studying a scientific community, including operations such as filtering, time slicing, pruning, merging, and visual inspection.

Finally, another approach to studying scientific publications focuses on the aspects

of time; examples include research fronts analysis (Morris, Yen, Wu, & Asnake, 2003) and historiographs (Garfield, 1973). Although this article focuses on summary graphs of authors, articles and conferences throughout their history, it presents a few time-related aspects as well.

Research Methods

The two primary components of this work were data collection, cleaning and processing followed by visual exploration of the resulting datasets. In fact, these occurred in numerous stages and cycles. Often it was the visual exploration that revealed faults with the data cleaning or suggested new data to collect or combinations and calculations that would be useful to explore.

Data Collection and Processing

We restricted our analysis to the four conferences CHI, UIST, AVI and InfoVis for a variety of practical reasons. First, the Metadata of the first three is managed by ACM, is publicly available in a usable format and is relatively complete and accurate compared with that from other sources.

For example, IEEE Digital Library metadata does not contain reference and citation information. Since this information has been added up to 2003 by the IEEE InfoVis 2004 Contest organizers, we have been able to use it. In contrast, the HCI Bibliography (`hcibib.org`) does not provide references and citations so we have not used it.

Another consideration was limiting the dataset size, which is already near the limit of what many current visualization tools can analyze. We also considered the selected conferences as a good overview of the HCI field. In particular, while data from the ACM Computer-Supported Cooperative Work (CSCW) conference would have been interesting to include, we opted not to because two analyzes of this community have been published, one in 2004 and another in 2006 (Horn et al., 2004; Jacovi et al., 2006). Finally, we

restricted our dataset to conference data because they are considered as the most important form of publications by HCI practitioners. Furthermore, journal articles and books are sufficiently different in their time scale and impact on the community that we felt comparisons between the two would be difficult.

While it may be argued that the AVI conference is insignificant in comparison to the other conferences selected for this analysis, we picked it due to precisely this reason: it is a young and upcoming conference which exhibits many of the typical patterns of newcomers. The analysis shows many of these signs of a still-immature conference, such as unstable co-authorship network and unformed communities.

Data Collection.

We began with the InfoVis 2004 Contest dataset, which covers the InfoVis conferences from 1995 to 2002. The data originally provided by the IEEE Digital Library (DL) had been extensively cleaned and corrected by the contest organizers. We used a version with additional curation provided by the University of Indiana as part of their contest submission. The datasets for the other 3 conferences were provided by the ACM Digital Library: the CHI conferences from 1983 to 2006, the UIST conferences from 1988 to 2005, and the AVI conferences from 1994 to 2006 (AVI is held every 2 years). The ACM DL provided an XML file for each conference with the title, authors, and other information about each article, including the unambiguous ACM identifiers of the articles it references wherever the curators were able to resolve them (see Figure 1).

Figure 2 shows an overview of the timeline of the four conferences as well as the coverage of the publication data used in this article. Note that data is missing for AVI 2002 and that the coverage of InfoVis ends in 2002.

We only collected information for full-length papers, excluding short articles, poster and demo submissions, contest entries, keynotes, panels, and so forth. For each conference, we collected the following information: proceedings ACM identifier, conference

ACM identifier and its acronym, proceedings title, proceedings description and copyright year. For each article, we collected the following information: article ACM identifier, title, subtitle, list of keywords attributed by the authors, abstract, page numbers in the proceedings, a list of citations to the article with the citing paper's ACM identifiers where identified, a list of authors, and their authoring sequence number. Self-citations were not removed from the dataset. Finally, for each author we collected their ACM identifier, first, middle and last names.

Data Processing.

It is important to note that our dataset is incomplete. First, the ACM metadata is incomplete, especially for early conferences. While it does contain basic information such as title, authors, and dates for each conference article, not all references are present, and not all references that are present have been unambiguously resolved. Secondly, because we only processed files from the four conference series, even identified articles from other conferences have missing detailed information, such as authors. Because such missing data could easily have misled our analysis, considerable caution is advised in interpreting both the visualizations and the statistics.

In addition to missing information, the datasets contain duplicated author identifiers, a common problem when dealing with publication data. Author names may be misspelled or use initials instead of full names, or authors may change their names or use different combinations of formal and informal names and initials on different papers, producing multiple identifiers we call *aliases* for a single person. Our efforts were aided by the recently-developed *D-Dupe* program from the University of Maryland (Bilgic, Licamele, Getoor, & Shneiderman, 2006). D-Dupe uses both name and co-authorship similarity in an interactive process to resolve aliases. We divided our de-duplication process into four stages, from the easiest to the more complex cases.

- We merged authors according to an alias attribute previously computed for the

InfoVis 2004 Contest. Katy Börner and her students had cleaned this dataset manually. For each of the 109 authors with aliases, they added an attribute to the original identifier in their database.

• We merged authors with exact similarity of last, middle and first names. Authors who used only a last name and a first name were merged them according to 2 criteria: if they had at least one co-author in common, and if their name subjectively and/or objectively did not seem to be common. (For example, two "Pedro Szekely"s would have been merged, but not two "J. Smith"s.) To define if a name was common or not, we used our own knowledge in addition to the search feature of D-Dupe. In the above example, for instance, a D-Dupe search on "Szekely" returns only 4 results, against 39 for "Smith".

• We merged authors with similar last name and more than one co-author in common. In that case we also used our knowledge of the field to avoid merging, for example, husband and wife Gary M. Olson and Judith S. Olson who have 7 co-authors in common. Still, we merged the 7 identifiers of William Buxton (as W. Buxton, William Buxton twice, William A. S. Buxton, Bill Buxton twice and B. Buxton).

• Finally, we had to deal with more complex cases: two persons with similar last names (relatively common) without any co-authors in common. To solve that case, we searched for information on the Web, looking for home pages and list of publications. Interestingly, in these cases the results were almost equally divided: half turned out to be the same individual collaborating with different teams, and half were different persons. This result implies that such cases will be difficult to resolve automatically.

The process took almost a day. We stopped when name similarity was less than 80%, being aware that duplicated authors still remained. We found a total of 516 aliases over the 6143 authors (8.3%). The maximum number of aliases was 7 apiece for Ben Shneiderman and William Buxton.

Visual Exploration Method

The collected results from the above data collection and processing produced a graph with 26,942 vertices and 118,865 relations. This graph contains three types of vertices: 332 conferences, 5,109 authors and 21,501 articles. Of the articles, 18,573 are missing some information, and 4,797 do not even have an ACM identifier. The network has three types of relations: 3,254 edges linking articles to the conference they appeared in, 9,030 edges linking articles to their authors, and 85,319 edges between articles (i.e. references). From these three, we computed additional relations: author-author for both co-authorship (10,631 relations) and citation, and conference impact (citations aggregated at the conference-conference level).

As stated in the introduction, we used an exploratory process to analyze the cleaned HCI publication data. This process does not require a priori hypothesis or questions to evaluate, but seeks to generate and evaluate hypotheses—about global and local trends and outliers—interactively during the exploration.

Visualizing and interacting with this data requires a system able to handle large graphs. Our analysis primarily used MatrixExplorer (Henry & Fekete, 2006) and NodeTrix (Henry, Fekete, & McGuffin, 2007) (both built upon the InfoVis Toolkit (Fekete, 2004)), GUESS (Adar, 2006) (based on JUNG[1]), and the R statistical package (R Development Core Team, 2006).

We used GUESS and its powerful scripting language to query graphs and manipulate their attributes. However, handling these large node-link diagrams induced some delay. Getting a readable overview of the full graph was also a challenge. For this reason, unlike most other studies, we choose to use an adjacency matrix representation of the graphs to explore the data in ways that would have been difficult otherwise.

We used the MatrixExplorer and NodeTrix tools to provide us with both matrix and node-link representations of the graphs. These systems offer interactive tools to

manipulate matrices (filtering, ordering and visual variable affectations) and allows for synchronized node-link diagrams. They also suffer some delay handling the full graph (especially to compute reordering), but the readability of the final representations was far better than with a node-link diagram.

We used matrix representations to explore the graph following a cyclic exploration process we will attempt to describe. We loaded our full dataset and filtered it by types of vertices, group of conferences and/or type of relations. For example, we extracted the co-authorship networks for InfoVis conferences, the citations network across conferences, or the citations network of CHI authors. For each of the filtered graphs, we then visualized its macro-structure: the connected components size and number followed by the analysis of each component independently. For each component, we interactively applied reordering, filtering, and visual variable affectations. We ended up with a set of insights such as communities or patterns for each filtered networks. At this stage, we created node-link visualizations of filtered graphs for each insight we found interesting. We fine-tuned the node-link visualizations in turn to get readable representations illustrating our findings.

At each stage, our analysis raised many additional questions. Organizing the exploration process to avoid diverging in several directions was difficult; since we were tempted to follow each insight independently. We recorded all the interesting questions but attempted to explore in a breadth-first manner instead of analyzing every individual question in depth, which often would have required time-consuming investigation on the Web or interviewing experts.

Although adjacency matrices were effective for exploration, presenting them on a static page with limited space is a challenge. Therefore, we present both zoomed views of our large matrices and node-link diagrams of filtered networks to illustrate our analyzes.

Results

This section describes the results of our visual exploration process. It primarily documents many observations, tentative explanations and questions for further analysis.

Overview

The first few subsections that follow present fundamental components of the HCI field and our datasets: its highly-cited authors and articles, the general characteristics of the four major conferences (CHI, UIST, AVI and InfoVis), and also an analysis of the evolution of their topics over the years.

Our relatively simple data analysis of this data, using primarily simple statistics, histograms and plots, explained many general characteristics of the data, but it also raised many additional interesting questions. We present a subset of these additional results we actually explored, and also try to give a feeling for a variety of additional queries that can be performed by filtering, combining, and correlating the data.

The last two subsections are a more in-depth analysis of two networks derived from the original data: citation networks for conferences, articles and authors, and co-authorship networks between researchers. Together, they provide a wealth of data about the structure of the HCI community: the influence of different researchers, institutions and conferences; the groups of researchers who collaborate strongly and the wider-ranging collaborations between them.

Authors

We used three measures to identify important researchers of the field (Figure 3). We collected the total number of articles accepted to define the *most prolific authors*. We computed the number of citations to researchers' articles to define the *most cited researchers*. Finally, we computed the social network analysis measure of *betweenness*

centrality for each researcher in the largest connected component of the co-authorship networks for each conference and for all the conferences together. This measure is an attempt to determine how central an actor is by counting the number of shortest paths between other authors that go via this researcher.

The common social-network concept of "betweenness-centrality" in this context must be interpreted carefully: it may not necessarily indicate success. For example, researchers who move from one institution to another or students who graduate and take a job elsewhere become more central not because of their work *per se*, but because of geographic (topographic) factors. Nevertheless, very central actors do link communities and are therefore perceived as central.

Citations and Number of Articles.

When examining Figure 3 and the general statistics on authors, we observe a correlation between the number of citations and the number of articles. In general, the most cited researchers are also the most prolific, implying that they are actively contributing to the field in terms of quality and quantity. The five most-cited include the trio of Stuart Card, Jock Mackinlay and George Robertson (abbreviated as Card-Mackinlay-Robertson), followed by William Buxton and Ben Shneiderman.

We notice two exceptions to this trend: Edward Tufte and Ravin Balakrishnan. Edward Tufte has only two referenced works (both books), but he is cited almost forty times. This is easily explained: Tufte has few publications in this field because he is not an HCI researcher, but these books are seminal works for information visualization that are frequently cited by articles in the field. Ravin Balakrishnan is exceptional in the opposite direction: the sixth most prolific author with almost forty published articles, he is nevertheless cited approximately 50% less than similarly-prolific authors such as William Buxton or George Robertson. One interpretation might be that much of his work relies on specialized technologies unavailable to the majority of HCI researchers, which

limits the number of citations until and if they become more generally accessible. Another is that despite his high number of publications, he is much younger than the other most cited researchers and his articles did not had as much time to get cited.

Centrality.

Each conference has a different set of most-central researchers. For the CHI community, they are William Buxton, Thomas Landauer and Thomas Moran. For the UIST community, Scott Hudson is the most central researcher, while Takeo Igarashi, Ken Hinckley and Brad Myers have a similar betweenness-centrality. For InfoVis, Ben Shneiderman and Stuart Card are almost equal as the most-central figures. AVI has a very disconnected network with many small connected components, the largest of which contains only about twenty researchers. Therefore, we cannot rely on centrality measures to identify a particular researcher. Our conclusion is that AVI does not yet have a stable set of communities.

Considering the centrality of the aggregated conferences, notice that all the central authors of CHI, UIST and InfoVis are in the top twenty except Takeo Igarashi. This would imply that he does not collaborate much with the other central figures of HCI, and in fact he is more active in the interactive 3D community than in HCI. Figure 4 shows the collaboration between the twenty most central researchers in our dataset.

Articles

The two most cited articles across CHI, UIST, AVI and InfoVis are "Cone Trees: Animated 3D Visualizations of Hierarchical Information" (Robertson, Mackinlay, & Card, 1991), published at CHI in 1991 and cited 70 times, and "Generalized Fisheye Views" (Furnas, 1986), published at CHI in 1986 and cited 66 times (Figure 6).

Sources of Key Articles.

Articles from the CHI conference are the most heavily cited, representing six of the

top ten and seven of the top twenty. Interestingly, browsing the keywords of these articles reveals that the majority deal with information visualization. Moreover, Edward Tufte's book, "The Visual Display of Quantitative Information" (Tufte, 1983), one of the seminal works of information visualization, is the third most cited research work. While this shows that information visualization is an active topic in HCI, the result should be interpreted carefully; since visualization is the major focus of both the InfoVis and AVI conferences. Interestingly, articles from the InfoVis conference itself appear unexpectedly low in this ranking. The first, "Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents" (Wise et al., 1995), appears at the 20th position. These low impact numbers are probably partly due to the fact that information visualization as a specialized sub-field is more likely to cite general HCI papers than the reverse. However, the ages of the conferences are another key. Not only are authors likely to submit their best work to established conferences, but influential papers often amass citations for many years. Similarly, the first-ranked article of the AVI conference (held every other year since 1992 in Italy, but becoming much more prominent around 2000) appears only at the 43rd position: "Fishnet: a fisheye web browser with search term popouts" (Baudisch, Lee, & Hanna, 2004). By contrast, four articles from the also-small UIST conference appear in the top twenty, including one in the top ten: "SATIN: A Toolkit for Informal Ink-Based Applications" (Hong & Landay, 2000). Besides its longer history (at 18 years it is the second-oldest), this may also reflect UIST's more general HCI focus.

Another interesting insight is that two articles of SIGGRAPH 1993 are much-cited in HCI (in the 14th and 24th position): "Pad: an alternative approach to the computer interface" (Perlin & Fox, 1993). and "Toolglass and magic lenses: the see-through interface" (Bier, Stone, Pier, Buxton, & DeRose, 1993). This could suggest that SIGGRAPH has more impact on the community than internal conferences.

Authors of Key Articles.

Figure 5 shows references among main authors. Some key articles have a single author: George Furnas, Edward Tufte and Jock Mackinlay each individually authored one the field's ten most-cited articles. However, collaboration seems to be a more reliable route to success. Not only did the trio of Card-Mackinlay-Robertson co-author three articles in the top ten, but Jock Mackinlay holds the record of six articles in the top twenty, and Stuart Card is the single most-cited researcher in the field.

Conferences

For each paper, we extracted its number of *references* to other articles, and the number of *citations* from other articles to it. Then, for each conference we computed the number of articles accepted and the total numbers of references and citations for all its papers (Figure 8.) Conferences are grouped by category and ordered chronologically from the oldest to the most recent.

Accepted Articles.

A global trend for all four conferences is that the number of accepted articles has increased over the years. CHI accepted 60 articles for its first conference in 1983, rising to 151 long articles in 2006, a 2.5-fold increase over 23 years. AVI and InfoVis also slowly increased their number of accepted articles. UIST's pattern was more variable. On the average, it accepts about 30 articles. However, it started with 22 articles at its first conference, doubled the number of number of accepted articles in 1994; then remained almost stable with an average of 30 articles accepted each year. The only other exception was 2003, its 20th anniversary and the largest UIST conference, which accepted 50 articles. We observed that CHI 91, 92 and 93 accepted more articles than the following conferences: all three accepted over a hundred articles, around 30 articles more than in 1990 and 1994. One could ask if a particular event happened during these three following

years (1993 was the decennial of CHI), if the submitted articles were of better quality or simply if the program committee decided to increase the number of accepted articles.

Number of References.

As the number of accepted articles increased, obviously so did the total number of references. However, the average of references per article also increased. It was stable from 1983 to 1993 with 10 references per article (although the earlier conferences seem to have a high rate of missing references in the ACM Metadata) but increased to 15 references in 1994; then remained stable for 5 years before finally increasing in 1999 to 20 references and remaining stable through 2006. UIST 92 is the only exception with an average of 21 references per article. An interesting observation is that the average number of references evolved similarly for all conferences. Further investigation would be required to define if the number of pages of submitted articles increased or if another factor explains this increase.

Acceptance Rate and Most Cited Articles.

The CHI conference published its most-cited articles in 1986 (#1 most-cited), 1991 (#2, 4 and 5), 1997 (#8) and 1994 (#9). However, Figure 7 shows that the conference's acceptance rates in those years were relatively high: 39% in 1986 (the highest ever), 23% in 1991, 24% in 1997 and 27% in 1994—versus its historic average, the lowest being a 15% acceptance in 2002. Typically, a low acceptance rate is an indicator of quality: there must be an abundance of strong work if so many papers are rejected. However these results do not concur. Does a low acceptance rate imply a more conservative article selection process that deters or filters out unconventional, ground-breaking articles?

Keywords

Our data contains information about the additional keywords authors have added to their articles (i.e. beyond the standardized ACM Computing Classification System[2]

keywords required for some conferences). These keywords are interesting because they serve as indicators to the ideas and concepts that were current in the scientific communities at different points in time.

Figure 9 shows a frequency visualization of the 100 most common terms in the combined keyword corpus for all conferences in the dataset (4,843 unique keywords in total). Here, keywords are scaled in size according to their relative frequency of appearance in the dataset. Looking at this figure, it is clear that "information visualization" (95 counts) is a key concept in the HCI community, but that terms like "CSCW" (62 counts), "ubiquitous computing" (57 counts), and "visualization" (52 counts) are important as well.

In Figure 10, we see similar frequency visualizations for the 50 most common terms of the individual conferences. We notice that the CHI conference (3,321 terms) has a much wider variety of terms than any of the other three conferences, and it is clear that CHI has a broader scope than the others. Also, the emphasis on information visualization is less pronounced for the CHI dataset, and the most common term here is actually "CSCW" (46 terms as opposed to 38 for "information visualization"). Both AVI (494 terms) and InfoVis (474 terms) are much more focused on visualization. Looking more closely at the individual keywords it seems that AVI has a wider array of general HCI subjects, whereas InfoVis—not surprisingly—focuses on visual representations of different kinds of data. Finally, the UIST (1,206 terms) conference shows a mix of the other three, yet has also a strong emphasis on user interfaces, toolkits, and programming.

Finally, we are also interested in studying the use of these keywords and concepts over time to get an idea of how ideas and trends rise and fall in the history of the four conferences. Figure 11 presents a timeline from 1983 to 2006 of the 59 most common keywords for all conferences. Darkness indicates high counts, so we can immediately notice the high emphasis on information visualization and interaction techniques in 2000.

Other insights include the introduction of the term information visualization in 1991 (corresponding to the three big papers published by PARC at CHI that year (Card, Robertson, & Mackinlay, 1991; Mackinlay, Robertson, & Card, 1991; Robertson et al., 1991)), the large number of popular concepts that were introduced in 1992, and the late shift to trends such as privacy, ethnography, and, particularly, ubiquitous computing in the 90s.

Of equal interest are keywords that no longer are in use, or which have exhibited periods of revival. For the former category, "user interface management systems" is a good example, appearing only in articles published in 1987 and then never again. The term "constraints", similarly, appeared in 1992 and then immediately went out of fashion. For the latter category, the term "usability" is perhaps the best example. It appeared in the very first CHI conference in 1983; then disappeared; made a strong comeback in 1992, remained prominent for a long time, but has not been seen since 2004.

Citation Networks

This section analyzes three citation networks: citations between conferences, between articles and between authors. Conference citations show the impact of each conference on the others; article citations highlight key articles and their relationships. The author citation network has the most interesting patterns, because how authors cite each other reveals patterns in the community. Citation patterns reveal many influences, and demonstrate research trends over time.

Citations Between Conferences.

Figure 12a is a matrix visualization of the inter-conference citation network, showing how the conferences reference each other. The four conferences, CHI, UIST, AVI and InfoVis, are arranged on the rows and columns, grouped by conference and then ordered by year, most-recent first. The darkness and numeric value in each matrix cell

show the number of citations *from* the conference printed on the row *to* articles of the conference printed on the column. Elements on the diagonal are articles referencing another article in the same year, which are most interesting when they refer to articles submitted to the same conference.

Conference Impact   In informal interviews, researchers in the field frequently described the CHI conferences as having the most impact and prestige, pointing to its high number of articles published despite a low acceptance rate and large number of attendees as indicators that articles published at CHI have the most impact in the field. If we define the *impact* of a conference as its number of articles cited by other conferences over the years, we can observe that CHI conferences have indeed had a strong impact on the field. Figures 8c and 12a show that CHI conferences have a strong impact on the other three. Articles from CHI 99, CHI 97, CHI 95, CHI 92 and CHI 91 represent the majority of references, while CHI 86 has the unique distinction of having been referenced by *every* subsequent conference and year except UIST'03 and CHI'96. In terms of evolution across time, Figure 12a shows that a typical CHI conference has a high impact for the six or seven following years, whereas the impact of UIST or InfoVis is only high for three or four years.

Analyzing the impact of CHI conferences on AVI and InfoVis, we were interested to notice that only CHI 86, CHI 91, CHI 94 and CHI 95 have had a strong impact. To analyze this further, we visualized the impact of the CHI articles independently, filtering to keep only the most-cited ones, resulting in Figure 12b. Comparing the totals for articles with those for the whole conference brought an even more interesting observation: for at least two of the four high-impact years, virtually all the references from all the InfoVis conferences to a particular CHI conference year were to a single article. Fully 100% (42/42) of the InfoVis references to CHI 86 are for "Generalized Fisheye Views" (Furnas, 1986), and 85% (68/80) of the references to CHI 91 are for "Cone

Trees" (Robertson et al., 1991). It is surely significant that so much of the impact of the CHI conference on the InfoVis conference depends on these two early articles.

Average Number of Citations   Given that the impact (total citations) of a conference hinges significantly on a few very highly-cited papers, it is interesting to look at the average number of citations per paper in a conference as well. Interestingly enough, as Figure 8d shows, according to this metric it is UIST and not CHI papers that clearly have a higher average number of citations than the other conferences. At the other end, the smaller AVI conference, which usually has higher impact than the larger InfoVis, beats it even more dramatically in citations per paper.

UIST's higher average citation count comes at a price. Its number of accepted papers is one clue: UIST has accepted only 20-30 papers since the beginning of the conference, against nearly 120 for CHI 2006. This is possible because UIST has maintained a focus on core HCI topics, whereas CHI caters to a much wider range of interests and accepts papers on a broader range of topics. Like for InfoVis and AVI's focus on visualization (see below), these specialized topics may have a narrow audience and thus lower UIST's average impact. Clearly, UIST is more selective, but this may mean that its impact suffers.

It would be interesting to differentiate impact figures by sub-area, for instance by keyword. However, CHI's broader focus is also probably a reason for its larger total audience and impact.

Citation Patterns   Figure 12a also implies a correlation between the core topics of CHI and UIST. Although UIST is much smaller, almost every CHI conference has referenced at least one UIST article and vice versa, suggesting that the basic interests of their communities are strongly connected. Similarly, the two visualization-oriented conferences InfoVis and AVI cite one another. Interestingly, both conferences cite CHI and UIST

articles far more than the reverse. Presumably, this is a case of a specialized field needing to cite basic principles of the parent field (however note the above results about much of the impact depending on a few articles). It is also possible that CHI and UIST are less open to external articles.

Finally, an unexpected finding is an unusually high number of intra-citations (citations between articles within the same annual conference) for UIST conferences. The CHI 91 conference also shows a high number of intra-citations (33 articles referencing articles of the same conference year). Because intra-citations require authors to know of other submissions in advance, they indicate an intertwined community with many co-authorship relationships between groups, and/or prolific research groups that have multiple papers accepted in a year. By contrast, intra-citations are rare in InfoVis, which suggests that research groups there are less intertwined or individually less prolific than for CHI or UIST conferences. Alternate explanations might include reviewing styles and prejudices: for instance blind reviewing such as CHI uses would make it more difficult to "ration" multiple acceptances to the same research group.

Article Citation Network.

In an article citation network, articles are the vertices and references between articles are (directed) edges. We do not present any visualizations of article-citation structure as they are very large (up to 23000 nodes). Even if heavily filtered, they would be useless without readable node labels, which is difficult because article titles are typically longer than names. Therefore, the next few sections of this article present the results of interactive exploration, illustrated by selected highlights.

Structure An overview of the article citation network is useful to identify how articles in a conference reference each other, as well as articles outside. Unfortunately, it is impacted by missing data, in particular for article references outside our core datasets

that are much less effectively resolved.

A first observation is that for AVI and especially InfoVis, the graph of citations within the conference articles is much sparser than for CHI or UIST. CHI and UIST have a longer history, so one interpretation could simply that articles in these conferences have had more time to impact the field than articles at InfoVis and AVI. Another reason could be that CHI has far more articles in total (UIST does not, however), or that UIST and CHI generate more key articles.

Interesting observations concerning the citation matrix presented in Figure 12a is that CHI and UIST cite each other, AVI cites article from all three conferences, and InfoVis is more isolated, primarily citing articles in its own conference. Of the few links that point outside the InfoVis area (towards the top of the diagrams) in the UIST (right side) or CHI area (left middle and bottom part), most are to a very limited subset of articles, as previously discussed. This observation confirmed that a conference impact may rely on a small set of articles (Figure 12b).

Citation Patterns   The general observation is that most-cited articles reference each other. Within those, "Generalized Fisheye Views" (Furnas, 1986) is the only article cited by others without referencing any of the most cited—trivially explainable as it was written before them. This article is seminal in the history of both HCI and InfoVis, as its citations reveal. Studying the top twenty key articles, only two articles cite others without being cited by them: "The Table Lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information" (Rao & Card, 1994) and "Pad++: A Zoomable Graphical Interface System" (Bederson & Hollan, 1994). The explanation is also chronology: published in early 90's, they are the most recent of our most-cited article set.

Finally, we noticed that two of these articles cite one another: the "The Information

Visualizer: an Information Workspace" (Card et al., 1991) and "The Perspective Wall" (Mackinlay et al., 1991). Again, the explanation is trivial: both were written by the same authors, the trio of Card-Mackinlay-Robertson all then of PARC, and published at the same conference, CHI '91.

Author Citation Network.

In the author citation network, the authors are the vertices and their references to other authors are the edges. This network is derived from the article citation network by aggregating articles that connect citing to referenced authors. This network shows how the important contributors in the field influence each other.

Figure 13 presents heavily-filtered node-link diagrams of the author citation networks for CHI, UIST, InfoVis and AVI. Filtering all but the most-cited authors allowed us to see how they cite one another. Node size and darkness redundantly encode each researcher's total number of citations, while the width and darkness of the links do the same for the number of citations from one researcher to another.

Citation Patterns   A first observation is that the trio of Card-Mackinlay-Robertson appear prominently in both the CHI and InfoVis networks, referencing one another heavily in both article sets. An obvious interpretation was that they were referencing the breakthrough articles they co-authored in both HCI and information visualization.

In the CHI author citation network, we saw that CHI's single most-cited author, William Buxton, is heavily cited by six of the other leading researchers. All cite him much more than the reverse, with the striking exception of Abigail Sellen, whom he cites far more. He also cites Hiroshi Ishii and Scott Mackenzie relatively frequently.

Examining the InfoVis author citation network, we observed that Ben Shneiderman has a pattern similar to William Buxton. Curved links underlined the mutual citation of Ben Shneiderman and Christopher Ahlberg. These two collaborated (with Christopher

Williamson) on "Dynamic Queries for Information Visualization" (Ahlberg, Williamson, & Shneiderman, 1992), one of Ben Shneiderman's most-referenced articles.

Finally, the much smaller author citation networks of UIST and AVI did not show strong patterns of citations. For UIST, we could only observe that Scott Hudson is referenced most often by the most-cited authors.

Considering self-citation, we observed a global pattern that the most-cited researchers heavily reference their own work. This is not true for AVI, perhaps because many participants only began contributing after 2000; so the pattern has not had time to emerge (especially on a biennial schedule). The self citation trend is particularly strong for the Card-Mackinlay-Robertson trio at CHI and InfoVis, for Hiroshi Ishii and William Buxton at CHI, as well as for Ben Shneiderman at InfoVis and Scott Hudson at UIST. Our interpretation is that these authors of multiple breakthrough articles in the field naturally cite them.

Co-Authorship Networks

We analyzed co-authorship data in two stages. First, we surveyed the macro-structure of each conference community, describing its connected-components structure and global statistics (with some comparison to other fields. In the second stage, we performed a detailed analysis of communities we had identified within this data, first for the whole HCI community (aggregating the data of all four conferences), and then for each conference community independently.

Macro Structure.

A connected component is a maximal connected sub-graph: a vertex in one connected component has no path to any vertex from another connected component. In this context, this information told us whether the research field is primarily composed of distinct communities that do not publish together or a single one connected by various

degrees of co-authorship. Figure 14a is a bar chart of these connected components. Each bar represents *all* the components of a given size. Its height is the log of the component size, and the width represents the number of components of that size. Note that even at a log scale, CHI and UIST as well as the aggregated data of all the conferences show a single "giant component", a very tall and thin (because it has only one element) bar representing a component containing approximately half the authors, all of whom interact. This is shown more precisely in Table 14b. By contrast, the largest component in the InfoVis and AVI graphs is far smaller, representing only 13% and 9%, respectively, of their authors. The most likely explanation seemed to be that the citation patterns of these newer conferences had not developed as fully (as well as having time for students to graduate and researchers to move between institutions); so the joint publications that would link different community components have not had time to appear. Alternate explanations included commercial constraints in the visualization field (such as some research being done with very expensive hardware or proprietary software) that restrained collaboration between communities.

By way of comparison, Table 14c presents data on several fields extracted from (Newman, 2001) (Medicine, biology and computer science) and (Horn et al., 2004) (the HCI field). The HCI data in this table comes from a different source, HCIbib.org, which does not contain any information on article references. We computed similar measures for our own data, as (Table 14b) shows, to provide some comparison with other fields. However, these comparisons should be made with caution, for two reasons:

1. The percentage of incompleteness and errors in these datasets is unknown; and

2. Because the measures are computed on variables which often follow power-law distributions, averages might not be a good comparison.

Communities of HCI.

Our first analysis was performed on a network composed of the data of all four

conferences. Here, the largest component is a subgraph containing 2,522 authors. Standard node-link diagrams of such a large graph would be unreadable without heavy filtering. Instead, we used the adjacency matrix representation provided by our tool MatrixExplorer (Henry & Fekete, 2006). The analog of graph layout for this representation is *matrix reordering*: finding a 1-D ordering of the nodes that groups closely-related ones; so the patterns become visible. Traveling Salesman Problem (TSP) approximation algorithms give good results for reordering many kinds of data. By placing authors with similar co-authorship patterns nearby, ordering reveals community structures effectively (even preattentively) as blocks of adjacent edges.

Unfortunately, large matrix visualizations are even harder to fit on printed pages than node-link ones. Therefore, we present several *NodeTrix* visualizations of selected details of these graphs. This representation represents the large-scale network structure with a standard node-link diagram but converts dense regions that would be unreadable in node-link as multiple small matrix representations. It includes flexible tools for dragging and dropping groups of nodes from one to the other. The NodeTrix visualization is particularly effective for small-world networks. For co-authorship networks, strongly-connected communities appear as preattentively-visible block patterns on the matrix display. We created NodeTrix representations by interactively dragging visual clusters appearing in a matrix representation into a NodeTrix visualization window. Very large clusters were edited into separate communities to show their detailed structure. This visualization allowed us to represent the main communities together with the details of their connections. However, because of the interactive editing and labeling, the results are subject to interpretation.

Figure 15 presents the visualization created during our analysis process. Reordering the matrix of the largest component of the co-authorship network reveals several visual clusters that we have outlined in the upper right corner. A *visual cluster* in the matrix is

a sub-matrix denser than the others. It means that the researchers of this sub-matrix collaborate with each other, i.e. form a *community*. By zooming in to examine these clusters closely and applying our own knowledge of the domain, we discovered that these clusters group researchers primarily by institution or by research topic.

Dragging these visual clusters into a NodeTrix window and dividing them into smaller communities centered on a main researcher resulted in the visualization at the top of the Figure 15. A zoomed-in view in the lower left corner shows one of these communities in detail.

In the data combining all four conferences, we located four main communities:

• CMU-Toronto: a community centered on William Buxton that is composed primarily of researchers from Carnegie Mellon University and the University of Toronto;

• CSCW-UMD: a community of CSCW researchers that includes a large group of researchers from Nottingham University: Steve Benford and Chris Greenhalgh, and also researchers from other institutions such as Ben Bederson from the University or Maryland and Michel Beaudouin-Lafon from the University of Paris-Sud;

• PARC: a community centered on Stuart Card and Jock Mackinlay, containing Ben Shneiderman from University of Maryland as well as Elizabeth Mynatt from Georgia Tech;

• Microsoft Research: a community mainly centered on George Robertson, Ken Hinckley and Patrick Baudisch.

We broke these four large communities in smaller ones and present the NodeTrix visualization in Figure 15. Each small matrix is a community centered around a researcher and/or an institution. Two distinct patterns recur in these small matrices: crosses and blocks. Dark crosses indicate a single researcher who collaborates with many others, while dark blocks indicate groups of researchers collaborating with each other (a perfectly-collaborative block, meaning that each member interacts with every other member, is called a *clique*, which appears as a fully filled-in dark block; since there is an

edge in each position between them). For example, the detailed matrix view in the lower right corner shows Ken Hinckley is linked to many other researchers with a cross-pattern, while also being part of a smaller clique of Agrawala - Ramos - Hinckley - Baudisch - Robertson - Czerwinsky - Robbins - Tan. In NodeTrix, the links between the matrices show how communities are linked at a high level. The width of the link lines shows the number of researchers involved in the collaboration: for example, George Robertson collaborated with a third of the researchers in the PARC community and around half of the researchers in the Hinckley et al. community.

Interacting with the visualization revealed that Ben Shneiderman bridges the PARC and CSCW-UMD communities. He effectively collaborated with Stuart Card of PARC and also with researchers from his home institution, the University of Maryland, such as Ben Bederson and Catherine Plaisant. George Robertson is a bridge between Microsoft Research (his new institution) and PARC (his former one). The co-authorship collaboration patterns of other central researchers such as William Buxton have a more prominent cross pattern, showing that they are the center of collaborations with a large number of researchers. In the node-link regions between matrices, a cross pattern becomes a dense web of links converging on the central researcher.

The following sections describe these different communities in more detail. We present four zoomed-in visualizations of the largest component of the matrix. These show the clusters CMU-Toronto in Figure 16, CSCW-UMD in Figure 18 PARC in Figure 17 and a portion of the Microsoft Research community in Figure 15.

CMU-Toronto: The central researchers of this cluster are William Buxton, Thomas Moran, Brad Myers and Iroshi Ishii. Figure 16 is a matrix visualization showing the major part of this community centered on William Buxton. Shades inside the matrix mark the strength of the collaborations. Shades in rows and columns indicate the number of

citations of these researchers. It is clear that William Buxton has had many collaborations with the most-cited researchers. These researchers have collaborated with each other in small groups (noticeable as blocks in the matrix). For example, William Buxton, Ravin Balakrishnan, Tovi Grossman, Thomas Baudel, George Fitzmaurice and Gordon Kurtenbach form a near-perfect clique. Thomas Moran and Brad Myers appear here as collaborators of William Buxton, but the remainder of the communities formed around these two individuals are located off-axis, in another part of the matrix that is not shown. Finally, the community centered on Iroshi Ishii is visible at the upper left corner of the matrix. His pattern is similar to William Buxton, a large "cross" of coauthors who did not collaborate strongly with one another.

CSCW and UMD   Figure 18 shows two large cliques connected through Ben Bederson as well as a large community centered on Chris Greenhalgh and Steven Benford (sparse block occupying the main part of the matrix). The community at the upper left mainly contains researchers from the University of Maryland linked to Steven Benford. The second large block connects members of the European Union-sponsored InterLiving project. It is interesting to note that the strongest collaboration of this community is Benford-Greenhalgh (11 co-authored articles) and that they both have very similar connection patterns, i.e. they have collaborated with the same researchers. The community centered on them can be further broken down into several smaller groups (blocks) of researchers who collaborating actively with each other.

Microsoft Research   An enlarged NodeTrix view of this community appears in the lower left corner of Figure 15. The NodeTrix view of its detailed structure includes three main sub-communities labeled Baudisch et al., Robertson et al. and Hinckley et al.). A general observation for this cluster is the strong collaborations within Microsoft Research, especially between George Robertson and Mary Czerwinski who co-authored 16 articles.

This strength is visible in the matrix representation as gray-scale indicates the strength of the collaboration.

PARC: The NodeTrix representation of this community has wide links going to George Robertson, and also to the Berkeley community, Alison Woodruff in particular. Figure 17 is a zoomed-in view of the matrix showing the Alison Woodruff and Keith Edwards community. It shows small sub-communities, such as the one centered on Peter Pirolli connected to Stuart Card and Jock Mackinlay, the one centered on Alexander Aiken connected to Alison Woodruff and the one centered on Elizabeth Mynatt, connected to Keith Edwards. Ben Shneiderman also appears in this community, primarily because of a single reference, the much-cited handbook "Readings in Information Visualization" he coauthored with Stuart Card and Jock Mackinlay.

UMD-InfoVis: We did not break out this community as a separate chart, but we annotated it off-axis in the original matrix. Several well-known InfoVis researchers appear in this community: Tamara Munzner(British Columbia), Martin Wattenberg(IBM) and Ben Shneiderman's collaborators Christopher Ahlberg and Christopher Williamson. This is easily explainable as an artifact of our reordering algorithm, which places the largest groups in the center of the matrix as it computes a 1D ordering. Because of Ben Shneiderman's surprising appearance in the PARC cluster in the primary ordering, the remainder of this community of which he is the center was pushed to the side of the matrix, still intersecting with him but off-axis. Note that Ben's cross pattern therefore appears as separate vertical and horizontal pieces in the symmetrical upper and lower matrices.

Communities of Each Conference. This section presents NodeTrix visualizations for the CHI, UIST, InfoVis and AVI conferences separately, attempting to show both

communities and important actors.

As we zoom into the NodeTrix visualization, the rows and columns of each matrix become readable, and thick consolidated links resolve into specific links between individual researchers. The figures do not provide detailed view of the whole networks here because of the lack of space, but they show a few selective enlarged portions. However, it must be kept in mind that we performed editing, analysis and labeling using interactions on the representation (drag and dropping elements to and from matrices) and zooming to produce these representations.

CHI: The organization of the co-authorship network containing only CHI data is shown as a NodeTrix in Figure 19a. The matrix visualization of the whole largest component revealed a main visual cluster centered around William Buxton and Thomas Moran. We present a zoomed-in view of the matrix visualization showing this cluster in Figure 19b.

By interactively filtering and ordering the matrix visualization of the largest component, we were able to distinguish five different communities (Figure 19b):

1. The largest community centered on William Buxton and Thomas Moran, including Abigail Sellen, William Gaver, Paul Dourish and Shumin Zhai. We also notice that a smaller community formed around Hiroshi Ishii;

2. The Brad Myers and Stuart Card community;

3. The community centered on Steve Benford and Chris Greenhalgh

4. The community centered on Ravin Balakrishnan and Ken Hinckley; and

5. The CMU community centered on Scott Hudson, Sara Kiesler and Robert Kraut.

Other zoomed views in the co-author matrix show interesting communities such as a clique (fully connected community) formed by researchers of UMD and and the French INRIA research institute, or the Microsoft Research community where collaboration between researchers is strong (9 articles co-authored by Mary Czwerwinski and George

Robertson).

It is interesting to note that the largest community in the NodeTrix visualization above appears to be the one centered on Steven Benford and Chris Greenhalgh, but this is only because we split up William Buxton's community into several smaller ones. This breakdown was natural, because Buxton's matrix has many links to other matrices. This indicates that William Buxton's many collaborators are actually active in many small communities, but all these communities are pulled into Buxton's community by their central members who collaborate with him, just as Ben Shneiderman's UMD community was dragged beside PARC. These strong effects of a few individuals on the ordering may not be optimal for showing each group's individual structure, but they do outline the largest communities clearly. This is evident in the zoomed-in matrix view in Figure 19b, which shows almost all the collaborators of William Buxton in a single clearly-delineated view.

UIST: Figure 20 shows the largest component of the co-authorship network of UIST as a NodeTrix visualization. Two sections have been enlarged to show several communities in details.

First, central actors are identifiable because their large number of connections and often make them bridges between communities. We can identify Ken Hinckley, Ravin Balakrishnan, Elizabeth Mynatt, Scott Hudson and Keith Edwards as central actors in UIST. It is interesting to notice that Elizabeth Mynatt is a bridge between the community centered on Blair MacIntyre and the rest of the network. Similarly, Igarashi acts as a bridge between researchers from University of Tokyo and the community centered on Jun Rekimoto.

As before, the cross and block patterns indicate the extremes of collaboration via a single individual and widespread collaboration between many members. In a node-link

diagram, the cross becomes a star pattern: the others collaborate often with the center actor but rarely with one other. Usually, this can be interpreted as a senior researcher advising junior ones. In Figure 20, we can identify these types of communities centered on Ravin Balakrishnan, Gordon Kurtenbach, Scott Hudson, and Keith Edwards and Jun Rekimoto.

The zoomed-in matrix in the lower left corner of this figure shows the largest community centered on Scott Hudson and Keith Edwards. In this community, we can notice that collaborators of Keith Edwards tend to collaborate with each other, as shown by the three blocks in the upper left corner of the matrix. Other examples of this pattern can be found in two matrices labeled PARC as well as in the community centered on Ken Hinckley: Microsoft Research, and the community labeled Berkeley. We characterize this as a mixed pattern, with a dark cross centered on one researcher, but included in a fairly dense block of mutual collaboration. As we previously saw for Ken Hinckley, the block refers to the strong connections within Microsoft Research: the cross is composed of researchers who only collaborate with Hinckley.

The zoom on the lower right corner clearly shows the two patterns. Ravin Balakrishnan has a high number of collaborators who did not collaborate with each other, whereas Forlines in the upper matrix is a bridge between two cliques of researchers who collaborate extensively with each other.

InfoVis: Figure 21 shows the largest component of the co-authorship network of the InfoVis conference. The lower right corner shows the overview of whole InfoVis matrix, labeling the main actors of this network: PARC and Ben Shneiderman. The largest cross identifiable is Ben, the most central actor in the InfoVis community.

The NodeTrix representation in the lower left corner shows how Ben Shneiderman acts as a bridge to the other UMD researchers grouped in a community centered on Ben

Bederson.

Finally, the upper part of the figure is a zoomed-in NodeTrix view showing how the PARC community collaborates with other communities. It is interesting to note that Berkeley and Microsoft Research strongly collaborate with each other. Similarly Stuart Card, Jock Mackinlay and Ed Chi collaborators are strongly connected.

AVI: Because the co-authorship network of AVI is quite small, we were able to fit the full matrix representation in Figure 22. This matrix is composed of many connected component, identifiable as disconnected blocks placed on the matrix diagonal. We present the details of several of these blocks as NodeTrix visualizations above and below the diagonal. The NodeTrix view of the largest component displayed in the bottom left of the picture shows that Patrick Baudisch from Microsoft Research is the central researcher of this component. The zoomed-in view on the upper right side of the matrix shows the connected component containing the most-cited researcher within AVI: Michel Beaudouin-Lafon from the University of Paris-Sud.

The collaboration within AVI must be interpreted with caution, because the conference has only become prominent since 2000 and is held only biannually (and also because the 2002 data is missing). However, these features make this conference data an interesting contrast to the others: a co-authorship network at a very different state of maturity. Relative to CHI or UIST, its network is very disconnected and with very low collaboration strength; since most research groups have only submitted a limited number of articles here. It is interesting to note that this network still presents a small-world effect, however.

Author-Author Collaboration. Finally, in Figure 23, we present node-link diagrams of the co-authorship networks filtered by number of citations. The node darkness represents the researchers' number of citations, and the node size their total number of

articles published. The darkness and width of the links redundantly encode the strength of the collaboration, i.e. the number of co-authored articles.

These four node-link diagrams reveal how most cited authors collaborate with each other. They highlight once again the three researchers Card-Mackinlay-Robertson who collaborate in both the CHI and InfoVis communities.

The global trend is that the most cited-researchers are both the most prolific and also have the largest number of collaborators. For all the conferences, most co-authors collaborate with each other. Within CHI and UIST, we observe that these collaborations are strong and shaped as a star pattern centered on the most cited authors: William Buxton and Scott Hudson, who have a large number of co-authors, but these co-authors do not collaborate strongly together.

Within InfoVis and AVI, the most-cited authors also have a high number of collaborators. The pattern of collaboration of InfoVis is different from a single star shape: the collaboration seems more distributed, which makes sense given the relatively fragmented connected-component structure seen in Figure 14a.

Insights and Interpretation

In this section we try to interpret and summarize the results we collected during the analysis process.

Strategies to Produce Key Articles

In light of our data exploration, we identified several different "strategies" that the most-cited researchers (authors of key articles) could be said to follow.

Have the Right Idea at the Right Time Write a book or an article in an emerging field. For example, Edward Tufte's *The Visual Display of Quantitative Information* (Tufte, 1983) presented key aspects of information visualizations just as

personal computers and spreadsheets were giving a much larger group of people the ability to create them. A second example is George Furnas, who wrote his article on generalized fisheye views (Furnas, 1986) in the early years of the CHI conference.

Collaborate with Other Senior Researchers By working with other senior and respected members of a field, you can achieve much more than you can on your own. This strategy is clearly visible in Figure 4 where the collaboration Card-Mackinlay-Robertson emerges.

Supervise a Good Number of (Good) Students Work with your students to publish in few targeted conferences. This strategy is visible in the collaboration patterns of the key InfoVis researcher Ben Shneiderman (Figure 21) and the CHI key researcher—William Buxton (Figure 19a). The matrices in these Figures reveal large "crosses" for both of them, meaning that these authors have a high number of co-authors (students) who may not frequently collaborate with each other. As a bonus, if you chose and taught them well, and they become successful and prolific themselves, they may lift your numbers and connectivity even higher by collaborating with you. For example, the InfoVis section of Figure 13 shows the collaboration between Christopher Ahlberg and Ben Shneiderman.

Publish in the Right Conferences Select the venue for your papers wisely. The four conferences chosen for analysis in this paper are all well-regarded in the field; yet, there is a clear difference between their impact and average number of citations. The CHI conference remains the most prestigious of these, with the highest number of citations. However, UIST has a higher average number of citations per article, so it would appear that UIST holds a higher overall quality than all of the other conferences.

Collaboration Strategies

Whereas the previous publication strategies are based primarily on the researcher's own abilities, two more rely on collaboration. We identified two that depend strongly on the research environment. Co-authorship in private research institutions such as PARC or Microsoft Research has a very different pattern from that in universities such as University of Toronto or the University of Maryland. Researchers in the private institutions collaborate with one another more freely; so they appear in matrix representations such as (Figure 18) as blocks, showing that most of the researchers have co-authored several articles together. The appearance of university research group collaborations has a completely different pattern: each professor and senior researcher has a cross pattern showing their co-authorship with a large number of students they advise. The students rarely publish with one another or with outside researchers without including their professor. For example, Figure 16 shows William Buxton's collaborators. These different patterns suggest that senior researchers within university research group work on different topics or are in competition with each other, i.e. they relatively rarely collaborate directly with each other.

Our interpretation is that each of the above strategies is well-adapted for its institutional environment. In private institutions, researchers are judged by the number of citations and their quality so they collaborate to produce the best possible articles. In contrast, universities insist on clear delineation of each researcher's contribution for tenure, promotion and other rewards; the more individualistic strategy adopted by most professors is rational: the merit of each non-student author is clear even if the overall impact is less.

Ben Shneiderman

A major figure of the HCI community, University of Maryland professor Ben Shneiderman, applied an unusual mix of these strategies. He wrote reference books (which

we do not study), authored seminal articles in the main conferences and collaborated with most of the key researchers of the field. However, he collaborated with other senior researchers exceptionally often for a professor. He co-authored articles with the senior PARC trio of Card-Mackinlay-Robertson. However, his co-authorship pattern also shows he advised several students over years.

However, Ben Shneiderman never worked for a private research institute, where even more collaboration might have increased his impact. For example, while Stuart Card, Jock Mackinlay and George Robertson were productive on their own, they reached a critical mass of productivity when joining together at PARC. Furthermore, Ben Shneiderman built his own research group instead of joining an existing one, like William Buxton did in Toronto.

Invisible Researchers

The visualizations and statistics only show one part of the picture. Non-American research centers are almost invisible. Why are so few authors from European, Asian and South American research centers listed among the top researchers? This question requires investigations deeper than the scope of this article allows, but it should raise questions both for the selection process of the conferences and for the selection process of non-American research centers. Are conferences outside North-America being evaluated fairly? Is the review process of the CHI-UIST-InfoVis conferences strongly biased against non-native-English speaking researchers?

Conclusions and Future Work

This article presents our analysis and visualization of a selection of publication metadata of four major conferences in Human-Computer Interaction and Information Visualization: the ACM Conference on Human Factors in Computing Systems (CHI), the ACM Symposium on User Interface Software and Technology (UIST), the ACM Working

Conference on Advanced Visual Interfaces (AVI), and the IEEE Symposium on Information Visualization (InfoVis).

Instead of starting from a set of a priori questions, we relied on visual exploratory analysis. This paper shows the visualizations we used, and describes some of the insights we gleaned from them. We needed to use a breadth-first strategy because this form of investigation raised so many additional questions that an exhaustive analysis of each in turn was impractical. The results are presented as a combination of matrix and node-link representations of the publication graphs. Given the incompleteness and noisiness of the data, it is important to exercise caution when interpreting our results. Nevertheless, we believe these insights will be a good first step in documenting the history of HCI for the benefit of students, practitioners, and researchers alike.

This work took a somewhat unusual approach of performing visual exploratory data analysis on the data of a scholarly community, instead of the more common confirmatory approach of statistically evaluating its conformance with a model or a set of a priori questions. This paper shows a number of visualizations we used, and describes some of the insights we gleaned from them. What it does not describe are the many frustrations of performing this work with existing tools. No existing package for community analysis or graph drawing was adequate for more than a fraction of our needs. We needed to use a variety of tools and do considerable ad hoc custom programming; yet still many interesting questions could not be explored in the time available.

Another major frustration and limitation was the incompleteness of the data and the biases that may have been introduced by the selection of available data and the process of data cleaning (for instance, the result about number of references per paper appearing to rise in recent years for which more references can be resolved.) Fortunately, making digital library metadata complete and accurate for automated analysis has many benefits beyond studies such as this one; so the source data quality is likely to improve rapidly. Part of the

solution will be tools, such as the *D-Dupe* package that helped us resolve author identities, and literature mining tools being developed for bioinformatics and many other fields. These can resolve divergent author names and other inaccuracies in article citations with much less need for manual curation than ours required. At the same time, digital libraries and online resources will eliminate ambiguity closer to the source. Informal wiki-style repositories can use community editing may suffice. For definitive repositories such as digital libraries, authors could receive secure IDs allowing them to correct ambiguities in their own author and publication identifiers. Finally, when authoring tools, bibliography and article submission websites have authors of new papers select their citations from standardized lists, they will only need to verify that they have the correct reference once.

Improving the metadata quality will raise the quality of analyzes and visualizations. These will permit much deeper and more reliable understanding of what organizational, environmental or personal factors improve research, beyond the simple quantitative measures used today.

Acknowledgments

Color images of this article are available at
`www.lri.fr/~fekete/20YearsOf4HciConferences`.

References

Adar, E. (2006). GUESS: a language and interface for graph exploration. In *Proceedings of the ACM CHI 2006 Conference on Human Factors in Computing Systems* (pp. 791–800). New York, NY, USA: ACM Press.

Ahlberg, C., Williamson, C., Shneiderman, B. (1992). Dynamic queries for information exploration: An implementation and evaluation. In *Proceedings of the ACM CHI'92 Conference on Human Factors in Computing Systems* (pp. 619–626). New York, NY, USA: ACM Press.

Ahmed, A., Dwyer, T., Murray, C., Song, L., Wu, Y. X. (2004). WilmaScope graph visualization. In *Proceedings of the IEEE Symposium on Information Visualization.* Washington, DC, USA: IEEE Computer Society.

Baudisch, P., Lee, B., Hanna, L. (2004). Fishnet: a fisheye web browser with search term popouts. In *Proceedings of the ACM Conference on Advanced Visual Interfaces* (pp. 133–140). New York, NY, USA: ACM Press.

Bederson, B. B., Hollan, J. (1994). Pad++: A zooming graphical interface for exploring alternative interface physics. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (pp. 17–26). New York, NY, USA: ACM Press.

Bier, E. A., Stone, M. C., Pier, K., Buxton, W., DeRose, T. (1993). Toolglass and Magic Lenses: The see-through interface. In *Computer Graphics (SIGGRAPH '93 Proceedings)* (pp. 73–80). New York, NY, USA: ACM Press.

Bilgic, M., Licamele, L., Getoor, L., Shneiderman, B. (2006). D-Dupe: An interactive tool for entity resolution in social networks. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology* (pp. 43–50). New York, NY, USA: IEEE Press.

Börner, K., Chen, C., Boyack, K. (2003). Visualizing knowledge domains. In B. Cronin (Ed.), *Annual review of information science and technology* (Vol. 37, pp. 179–255).

American Society for Information Science and Technology.

Boyack, K. W., Klavans, R., Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351–374.

Boyack, K. W., Wylie, B. N., Davidson, G. S. (2002). Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology*, *53*(9), 764–774.

Card, S. K., Robertson, G. G., Mackinlay, J. D. (1991). The information visualizer, an information workspace. In *Proceedings of the ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 181–188). New York, NY, USA: ACM Press.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377.

Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E., Wylie, B. N. (1998). Knowledge mining with VxInsight: Discovery through interaction. *Journal of Intelligent Information Systems*, *11*(3), 259–285.

Egghe, L., Rousseau, R. (1990). *Introduction to informetrics.* Elsevier.

Fekete, J.-D. (2004, October). The InfoVis Toolkit. In *Proceedings of the IEEE Symposium on Information Visualization* (pp. 167–174). Austin, TX: IEEE Press.

Fekete, J.-D., Grinstein, G., Plaisant, C. (2004). *IEEE InfoVis 2004 Contest.* available at `www.cs.umd.edu/hcil/iv04contest`.

Furnas, G. W. (1986). Generalized fisheye views. In *Proceedings of the ACM CHI'86 Conference on Human Factors in Computing Systems* (pp. 16–23). New York, NY, USA: ACM Press.

Garfield, E. (1973). Historiographs, librarianship, and the history of science. *Toward a theory of librarianship*, 380–402.

Goffman, C. (1969). And what is your erdos number? *American Mathematical Monthly,*

76.

Grossman, J. W., Ion, P. D. F. (1995). On a portion of the well-known collaboration graph. *Congressus Numerantium, 108,* 120–131.

Henry, N., Fekete, J.-D. (2006). MatrixExplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics (IEEE Visualization Conference and IEEE Symposium on Information Visualization Proceedings 2006), 12*(5), 677–684.

Henry, N., Fekete, J.-D., McGuffin, M. J. (2007). NodeTrix: a hybrid visualization of social networks. In *Proceedings of the IEEE Conference on Information Visualization.*   (to appear)

Hong, J. I., Landay, J. A. (2000). SATIN: A toolkit for informal ink-based applications. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (pp. 63–72). New York, NY, USA: ACM Press.

Horn, D. B., Finholt, T. A., Birnholtz, J. P., Motwani, D., Jayaraman, S. (2004). Six degrees of Jonathan Grudin: a social network analysis of the evolution and impact of CSCW research. In *Proceedings of the 2004 ACM Conference on Computer-Supported Cooperative Work* (pp. 582–591). New York, NY, USA: ACM Press.

Jacovi, M., Soroka, V., Gilboa-Freedman, G., Ur, S., Shahar, E., Marmasse, N. (2006). The chasms of CSCW: a citation graph analysis of the CSCW conference. In *Proceedings of the 2006 Conference on Computer-Supported Cooperative Work* (pp. 289–298). New York, NY, USA: ACM Press.

Ke, W., Börner, K., Viswanath, L. (2004). Analysis and visualization of the IV 2004 contest dataset. In *Proceedings of the IEEE Symposium on Information Visualization.* Washington, DC, USA: IEEE Computer Society.

Kretschner, H. (1994). Coauthorship networks of invisible college and institutionalized

communities. *Scientometrics*, *30*, 363–369.

Lee, B., Czerwinski, M., Robertson, G., Bederson, B. B. (2004). Understanding eight years of infovis conferences using PaperLens. In *Proceedings of the IEEE Symposium on Information Visualization* (p. 216.3). Washington, DC, USA: IEEE Computer Society.

Mackinlay, J. D., Robertson, G. G., Card, S. K. (1991). The perspective wall: Detail and context smoothly integrated. In *Proceedings of the ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 173–179). New York, NY, USA: ACM Press.

Melin, G., Persson, O. (1996). Studying research collaboration using coauthorships. *Scientometrics*, *36*, 363–377.

Morris, S. A., Yen, G. G., Wu, Z., Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology*, *54*(5), 413–422.

Newman, M. (2001). Who is the best connected scientist? a study of scientific coauthorship networks. *Phys. Rev.*, *64*, 016131–016132.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review, 45*, 167–256.

Newman, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences, 101*, 5200–5205.

Perlin, K., Fox, D. (1993). Pad: An alternative approach to the computer interface. In *Proceedings of Computer Graphics (SIGGRAPH 93)* (pp. 57–64). New York, NY, USA: ACM Press.

Price, D. (1965). Networks of scientific papers. *Science, 149*, 510–515.

R Development Core Team. (2006). *R: A language and environment for statistical computing.* Vienna, Austria. (ISBN 3-900051-07-0)

Rao, R., Card, S. K. (1994). The Table Lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proceedings of the ACM CHI'94 Conference on Human Factors in Computing Systems* (pp. 318–322). New York, NY, USA: ACM Press.

Robertson, G. G., Mackinlay, J. D., Card, S. K. (1991). Cone trees: Animated 3D visualizations of hierarchical information. In *Proceedings of the ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 189–194). New York, NY, USA: ACM Press.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, Connecticut: Graphics Press.

Watts, D., Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440—442.

Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., et al. (1995). Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of the IEEE Symposium on Information Visualization* (pp. 51–58). Washington, DC, USA: IEEE Computer Society.

Wong, P. C., Hetzler, B., Posse, C., Whitien, M., Havre, S., Cramer, N., et al. (2004). In-spire infovis 2004 contest entry. In *Proceedings of the IEEE Symposium on Information Visualization.* Washington, DC, USA: IEEE Computer Society.

Notes

[1] `http://jung.sourceforge.net`

[2] `http://www.acm.org/class/1998/`

Figure Captions

Figure 1. Resolved and unresolved references. References between the four conferences are resolved completely. Other references contained in the ACM DL are resolved with a unique identifier but no other information. References outside the ACM DL are not resolved.

figure.1 *Figure 2.* Timeline of the CHI, UIST, AVI and InfoVis conferences. The solid bars indicate the coverage of our publication data; AVI 2002 is missing.

figure.2 *Figure 3.* Statistics for authors and articles.

figure.3 *Figure 4.* Overviews of the HCI field in terms of collaboration (co-authorship). Each node represents a researcher with its size showing the number of articles published and its darkness represents the number of citations. Links represent co-authorship. Their width is the strength of these relations.

figure.4 *Figure 5.* Overviews of the HCI field in terms of influence (citations). Each node represents a researcher with its size showing the number of articles published and its darkness represents the number of citations. Links represent citations. Their width is the strength of these relations.

figure.5 *Figure 6.* Top 20 most referenced articles.

figure.6 *Figure 7.* Acceptance rate for CHI.

figure.7 *Figure 8.* Statistics per conference.

figure.8 *Figure 9.* Keyword frequency cloud for all four conferences (100 terms).

figure.9 *Figure (a).* Number of accepted articles

subfigure.8.1 *Figure (b).* Average number of references per article

subfigure.14.3 *Figure 15.* Largest component of the co-authorship for all conferences. We annotated the whole matrix with the different communities' labels (lower left corner), a zoom of the Microsoft Research cluster is provided on the lower right corner. Shades in the headers row and column indicate the number of citations. We dragged the visual clusters into a NodeTrix visualization, edit them and present the visualization in the upper part of the figure.

figure.15 *Figure 16.* Zoom on the main cluster: CMU-Toronto based upon the matrix of co-authorship for all conferences. In rows, areas are the number of articles a researcher published, in column the number of citations. Values in the matrix indicate number of articles published together.

figure.16 *Figure 17.* Zoom on a PARC community based upon the matrix of co-authorship for all conferences. In rows, areas are the number of articles a researcher published, in column the number of citations. Values in the matrix indicate number of articles published together.

figure.17 *Figure 18.* Zoom on a community CSCW - UMD based upon the matrix of co-authorship for all conferences. In rows, areas are the number of articles a researcher published, in column the number of citations. Values in the matrix indicate number of articles published together.

figure.18 *Figure 19.* CHI co-authorship network. Values in the matrix indicate number of articles published together.

figure.19 *Figure (a).* Overview of the CHI co-authorship network

subfigure.19.1 *Figure (b).* The largest CHI community centered on William Buxton and Thomas Moran

subfigure.19.2 *Figure 20.* UIST co-authorship network.

figure.20 *Figure 21.* The largest component of the co-authorship network of InfoVis. Communities are displayed as matrices.

figure.21 *Figure 22.* AVI co-authorship network is composed of many separate connected components. This figure shows the matrix of the complete network. Distinct connected components are visible in the matrix as non-connected blocks on the diagonal. Details of several of these components are shown in more details as NodeTrix representations with labels we consider representative. On the upper right of the matrix is the detailed component containing the most cited researcher in AVI. On the lower left of the matrix is the largest connected component.

figure.22 *Figure 23.* Co-authorship networks filtered by number of citations within the community. Nodes represent researchers: size shows the number of articles published to the conference, darkness shows the number of citations by articles of this conference. Links represent co-authorship, their width is the number of articles co-authored. These node-link diagrams use the LinLog layout with some manual modification to avoid label superposition.

figure.23

Most–Cited Authors

Stuart K. Card
George G. Robertson
Jock D. Mackinlay
William A. S. Buxton
Ben Shneiderman
Brad A. Myers
George W. Furnas
Benjamin B. Bederson
Hiroshi Ishii
Scott E. Hudson
Thomas P. Moran
Ramana Rao
Abigail J. Sellen
Maureen C. Stone
Peter Pirolli
Steven K. Feiner
Jun Rekimoto
Eric A. Bier
Ken Hinckley
Brygg Ullmer

□ Cited(max639)
■ Pubs(max47)

Most–Prolific Authors

Brad A. Myers
Stuart K. Card
Scott E. Hudson
Ben Shneiderman
William A. S. Buxton
Ravin Balakrishnan
George G. Robertson
Jock D. Mackinlay
James A. Landay
Peter Pirolli
Ken Hinckley
Shumin Zhai
Steven K. Feiner
Thomas P. Moran
Bonnie E. John
Dan R. Olsen
Elizabeth D. Mynatt
Benjamin B. Bederson
John M. Carroll
Mary P. Czerwinski

□ Cited(max639)
■ Pubs(max49)

Author Centrality
All Conferences

William A. S. Buxton
Brad A. Myers
Thomas P. Moran
Scott E. Hudson
Thomas K. Landauer
Stuart K. Card
Ben Shneiderman
Steve Benford
William W. Gaver
James A. Landay
Mary Beth Rossen
Dan R. Olsen
Bonnie E. John
Jock D. Mackinlay
Patrick Baudisch
Benjamin B. Bederson
George G. Robertson
Ravin Balakrishnan
Ken Hinckley
Hiroshi Ishii

InfoVis

Ben Shneiderman
Stuart K. Card
Stephen G. Eick
Jock D. Mackinlay
Mei C. Chuah
Benjamin B. Bederson
Chris Olston
Ed Huai–hsin Chi
S. F. Roth
Peter Pirolli
Steven F. Roth
Allison Woodruff
John Kolojejchick
George W. Furnas
Chris North
Nahum Gershon
Marti Hearst
James D. Hollan
Jade Goldstein
Michael Stonebraker

AVI

Patrick Baudisch
Maneesh Aarawala
Bongshin Lee
Mary Czerwinski
George Robertson
Desney S. Tan
Gonzalo Ramos
Ken Hinckley

CHI

William A. S. Buxton
Thomas K. Landauer
Thomas P. Moran
Scott E. Hudson
Stuart K. Card
Victoria Bellotti
James A. Landay
Ronald M. Baecker
Robert E. Kraut
Richard M. Young
Steve Benford
John M. Carroll
Bonnie E. John
Brad A. Myers
Abigail J. Sellen
Phil Barnard
George W. Furnas
Hiroshi Ishii
Benjamin B. Bederson
Bill Curtis

UIST

Scott E. Hudson
Takeo Igarashi
Ken Hinckley
Brad A. Myers
W. Keith Edwards
Ravin Balakrishnan
Thomas P. Moran
Patrick Chiu
Jock D. Mackinlay
Steven K. Feiner
Gene Golovchinsky
Elizabeth D. Mynatt
Laurent Denoue
Jonathan I. Helfman
Darren Leigh
Satoshi Matsuoka
Ian Smith
Dan R. Olsen
John F. Hughes
Gregory D. Abowd

Cone Trees: Animated 3D Visualizations of Hierarchical Information   CHI'91

Generalized Fisheye Views   CHI'86

The Visual Display of Quantitative Information   Book(86)

A focus+context technique based on hyperbolic geometry for visualizing large hierarchies   CHI'95

The information visualizer, an information workspace   CHI'91

Information visualization using 3D interactive animation   CACM'93

Tree–Maps: a space–filling approach to the visualization of hierarchical information structures   Vis'91

Visual information seeking: tight coupling of dynamic query filters with starfield displays   CHI'94

SATIN: A Toolkit for Informal Ink–Based Applications   UIST'00

The Table Lens   CHI'94

A review and taxonomy of distortion–oriented presentation techniques   TOCHI'94

Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics   UIST'94

Pad: an alternative approach to the computer interface   SG'93

Stretching the rubber sheet   UIST'93

Automating the design of graphical presentations of relational information   TOG'86

Brushing scatterplots   Techn'87

Spotlight: directing users' attention on large displays   CHI'05

Zliding: fluid zooming and sliding for high–precision parameter manipulation   UIST'05

Toolglass and magic lenses: the see–through interface   SG'93

Visualizing the non–visual: spatial analysis and interaction with information from text documents   IV'95

Citations to
top 20 HCI
Papers(max=70)

(a) Number of accepted articles

(b) Average number of references per article

(c) Number of citations per article

(d) Average number of citations per article

3d user interfaces agents animation annotation audio augmented reality awareness browsing children cognitive models collaboration computer mediated communication computer vision computer-mediated computer-mediated communication computer-supported cooperative work constraints context-aware computing CSCW data visualisation data visualization design process direct manipulation education educational applications email empirical studies end-user programming ethnography evaluation eye tracking field study fitts' law gestures goms graphical user interfaces groupware human-computer interaction hypertext information foraging information navigation information retrieval information scent

information visualization input device

input devices instant messaging intelligent interfaces interaction interaction design interaction techniques interactive techniques interface design internet java learning mobile computing mobile devices mobile phones multimedia multimodal interfaces navigation participatory design pda pen-based interfaces pointing privacy programming by demonstration programming by example prototyping sketching social computing speech recognition tangible interface task analysis text entry text input toolkits trust two-handed input two-handed interaction ubiquitous computing usability usability testing user interface user interface design user interface management systems user interface toolkits user interfaces user models user studies user study user-centered user-centered design video virtual environments virtual reality visual search visualization world wide web

AVI

ambient intelligence augmented reality clutter collaboration coordination cscw design guidelines design principles diagrams direct manipulation dynamic queries evaluation fisheye view flip zooming focus+context focus+context visualization gestures guides hierarchies information density information navigation information retrieval

information visualization

instructions instrumental interaction interaction model interaction objects interactive graphics layout mobile devices multimodal interfaces multiple views pen computing relational database software architecture tight coupling ubiquitous computing usability usability heuristics user studies user study virtual environments virtual reality visual interaction visual interface design visual interfaces visual placement visual techniques visualization zoomable interfaces

InfoVis

3d animation application domains approximate optimization data analysis data model data types data visualisation

data visualization decision tree design enhancements direct manipulation focus+context visualization graph drawing graph visualization graphical user interfaces graphics harmony browser hierarchies hierarchies. hyperbolic navigation hypertext iconic representations information analysis

information visualization

information visualization literature information-centric approach interactive systems interactive tree interactive visualization internet morphological analysis multiscale visualization multivariate data multiway tables pan and zoom parallel coordinates query processing semantic relationship text textual corpora treemaps user interface user interface design user interface environment user interfaces visage visualization world wide web vrml

UIST

3d graphics animation annotation augmented reality authoring collaboration computer-supported cooperative work constraints context-aware computing cscw direct manipulation distortion viewing dynamic queries end-user programming gestures graphical user interface graphical user interfaces groupware information navigation

information visualization input

devices interaction interaction techniques interface metaphors java mobile computing mobile phones multiscale interfaces navigation optimization pen-based interfaces programming by demonstration programming by example screen layout sketching speech user interfaces spreadsheets text entry toolkits two-handed input ubiquitous computing user interface user interface design user interface management systems user interface toolkits user interfaces window management virtual reality virtual worlds visualization

CHI

augmented reality awareness browsing children cognitive models collaboration computer-mediated communication computer-supported cooperative work CSCW design process direct manipulation education empirical studies end-user programming ethnography evaluation eye tracking fitts' law gestures goms graphical user interfaces human-computer interaction hypertext information retrieval information visualization input devices interaction design interaction techniques interface design learning mobile computing multimedia navigation participatory design privacy trust ubiquitous computing usability usability testing user interface user interface design user interfaces user models user study user-centered user-centered design video virtual reality visualization www

Term	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06

usability
user-centered
user interfaces
user study
human-computer interaction
graphical user interfaces
direct manipulation
gestures
input devices
user interface design
user interface management systems
user interface
interaction techniques
empirical studies
information visualization
participatory design
programming by demonstration
cognitive models
navigation
collaboration
interface design
multimedia
video
virtual reality
visualization
user-centered design
browsing
awareness
user models
constraints
goms
groupware
fitts' law
cscw
end-user programming
design process
education
hypertext
information retrieval
toolkits
animation
augmented reality
text entry
mobile computing
learning
computer-supported cooperative work
two-handed input
annotation
privacy
ethnography
children
ubiquitous computing
interaction design
data visualisation
world wide web
pointing
evaluation
eye tracking
computer-mediated communication

(a) Conference citations

(b) Conference impact

CHI

Shneiderman · Furnas · John · Minneman
Robertson · Card · Pirolli
MacKenzie · Kurtenbach · Zhai · Landay · Sellen · Moran · Bly · Whittaker
Myers · Buxton · Pedersen · Dourish · Tang · Gaver
Fitzmaurice
Druin · Ishii · Ullmer · Benford · Greenhalgh

UIST

Edwards · Green · Feiner · Robertson · Snibbe
Olsen · Linton · Mynatt · Mackinlay
Myers · MacIntyre · Rekimoto · Ishii · Hinckley
Cardelli · Smith · Chang · Bederson · Kurtenbach
Hudson
Henry · Hong · Landay · Guimbretiere · Winograd
Matsuoka · Tanaka · Saund · Moran · Van Melle · Stone

InfoVis

Conklin · Jurrus · Dam
Eades · Munzner · Ward · Brown · Becker
Hearst · Chi · Shneiderman · Cleveland
Reiss · Pirolli · Chalmers · Eick · Keim · Cox
Rao · Mackinlay
Lamping · Robertson · Feiner · Ahlberg · Johnson
Snibbe · Sarkar · Card · Roth · Tufte
Tversky · Foley · Kolojejchick
Hollan · Bederson · Furnas · Mattis · Bier · Stone · Fishkin · Goldstein · Williamson

AVI

Cox · Kramer · Woodruff · Shneiderman · Wang Baldonado
Vanderdonckt · Gillo · Smedley · Strothotte · Sonnet · North · Kuchinsky · Marchionini · Witkowski
Smith · Ramos · Haber
Carpendale · Livny · Bao · Wittenburg · Komlodi · de Bruijn · Ding
Lee · Ioannidis · Craft · Tse
Baudisch · Hinckley · Van Der Wege · Spence · Slaughter · Fawcett
Guimbretiere · Hanna · Beaudouin-Lafon · Card · LaLiberte · Ali-Ahmad
Mackay · Ellis · Dix · Butz
Pirolli

(a) Co-authorship connected components: size(log10) vs. number

	All 4	CHI	UIST	InfoVis	AVI
Number of authors	5 109	3 422	956	325	375
Number of articles	3 209	1 943	542	152	159
Articles per author	1.8	1.6	1.6	1.5	1.2
Authors per article	2.8	2.8	2.8	2.7	2.8
Average number of collaborators	4	4	3.8	3.2	2.9
Giant component	49%	50%	49%	13%	9%
Number of components	929	627	169	291	99

(b) Connected component count and size per conference

	Measure Biomed	HEP	CS	HCI
Number of authors	152 0251	56 627	11 994	23 624
Number of articles	216 3923	66 652	13 169	22 887
Articles per author	6.4	11.6	2.6	2.2
Authors per article	3.8	9.0	2.2	2.3
Average number of collaborators	18.1	173	3.6	3.7
Giant component	92.6%	88.7%	57.2%	51.3%
Mean distance	4.6	4.0	9.7	6.8

Woodruffe-Berkeley

Benford-Greenhalgh

Myers, et al.

Landay, et al.

Edwards

PARC

Bederson et al.

Buxton et al.

Baudisch et al.

Robertson et al.

Hudson et al.

Feiner et al.

Ishii et al.

Hinckley, et al.

Guimbretiere et al.

UMD-InfoVis

CMU-Toronto

Microsoft R.

Georgia Tech.

CSCW

Parc

Annotated matrix

Zoom on the matrix visual cluster for Microsoft Research

(a) Overview of the CHI co-authorship network

(b) The largest CHI community centered on William Buxton and Thomas
Moran

Berkeley

MacIntyre et al.   Tokyo   Rekimoto et al.

Mynatt et al.   PARC

Igarashi et al.   PARC

Hudson - Edwards et al.

Snibbe et al.

Pausch et al.   Hinckley et al.

Robbins et al.

Forlines   Balakrishnan et al.   Kurtenbach et al.

Dragicevick et al.

Forlines
Shipman
Ryall
Harsham
Yerazunis
Schmidt-Ni
Harada
Esenther
Miyachi
Wittenburg
Lanning

Balakrishn
Buxton
Fitzmaurice
Khan
Burtnyk
Bezerianos
Cao
Grossman
Wu
Malik
Ranjan
Beardsley
Raskar
van Baar
Vogel
Shen
Barnwell
Wigdor
Leigh

Lamping
Dourish
Thornton
Howell
Salisbury
Sediw
Newman
Izadi
Smith
Smith
Rodenstein
Marinacci
Rodriguez
Tanaka
Heiner
Newell
Henry
Yeatts
Stasko
Abowd
Mankoff
Bharat
Knudsen
Byrne
John
Fogarty
Forlizzi
Summet
Dietz
Lee
Hudson
Edwards
LaMarca

Toronto
Calgary MIT
Carpendale

Polytechnico Milan

GTE lab.

Germany

INRIA

	Beaudouin-Lafon	Zhai	Guiard	Bastin	Pasveer	Mackay	Blanch			Blanch	Mackay	Pasveer	Bastin	Guiard	Zhai	Beaudouin-Lafon
Beaudouin-Lafon									Beaudouin-Lafon
Zhai									Zhai
Guiard									Guiard
Bastin									Bastin
Pasveer									Pasveer
Mackay									Mackay
Blanch									Blanch

Uni. Roma Uni. Bari

MSR

	Baudisch	Robertson	Czerwinski	Tan	Agrawala	Ramos	Hinckley	Guimbretiere	Smith	Perer			Perer	Smith	Guimbretiere	Hinckley	Ramos	Agrawala	Tan	Czerwinski	Robertson	Baudisch
Baudisch												Baudisch
Robertson												Robertson
Czerwinski												Czerwinski
Tan												Tan
Agrawala												Agrawala
Ramos												Ramos
Hinckley												Hinckley
Guimbretiere												Guimbretiere
Smith												Smith
Perer												Perer

UMD

	Campbell	Parr	Bederson	Lee	Hanna			Hanna	Lee	Bederson	Parr	Campbell
Campbell								Campbell
Parr								Parr
Bederson								Bederson
Lee								Lee
Hanna								Hanna

MSR

	Smith	Meyers	Horvitz	Hutchings	Robbins			Robbins	Hutchings	Horvitz	Meyers	Smith
Smith								Smith
Meyers								Meyers
Horvitz								Horvitz
Hutchings								Hutchings
Robbins								Robbins

Stanfo

	Tversky	Hanrahan	Heiser	Phan			Phan	Heiser	Hanrahan	Tversky
Tversky							Tversky
Hanrahan							Hanrahan
Heiser							Heiser
Phan							Phan

CHI

Furnas
Erickson
Kellogg
Carroll
Myers
Schilit
Landay
Robertson
Mackinlay
Pirolli
Accot
Druin
Benford
Hudson
Hinckley
Czerwinski
Card
Pitkow
Whittaker
Zhai
Gaver
Balakrishnan
Pausch
Rao
John
Harrison
MacKenzie
Kurtenbach
Bellotti
Shneiderman
Grudin
O'Hara
Sellen
Buxton
Maclean
Young
Polson
Wharton
Ahlberg
Baecker
Fitzmaurice
Dourish
Lewis
Ullmer
Ishii
Moran
Pedersen
Nielsen
Mynatt
Baudel
Palen
Halasz
McCall
Bly
Minneman
Tang
Goldberg

UIST

Calder
Linton
Kurtenbach
Olsen
Robbins
Bederson
Green
Singh
Myers
Snibbe
Hong
Landay
Dannenberg
Balakrishnan
Sinclair
Czerwinski
Cardelli
Newell
Hinckley
Pausch
Robertson
Sukaviriya
Henry
Smith
Mackinlay
Ishii
Edwards
Igarashi
Chang
Guimbretiere
Stone
Tanaka
Moran
Chiu
Mynatt
Feiner
Matsuoka
Winograd
Van Melle
Saund
MacIntyre
Rekimoto

InfoVis

Bier
Goldstein
Mattis
Tufte
Brown
Sarkar
Tversky
Fishkin
Stone
Roth
Reiss
Becker
Eick
Card
Johnson
Snibbe
Cleveland
Mackinlay
Shneiderman
Williamson
Keim
Ward
Eades
Chi
Ahlberg
Feiner
Perlin
Pirolli
Robertson
Bederson
Furnas
Fox
Lamping
Rao
Munzner
Chalmers
Hollan
Foley

AVI

Hanna
Woodruff
Wang Baldonado
Lee
Shneiderman
Kuchinsky
LaLiberte
Baudisch
Dix
Ding
Marchionini
Ali-Ahmad
Slaughter
Wittenburg
Lanning
Beaudouin-Lafon
Pirolli
Komlodi
Tse
Mackay