

## College of Information Science and Technology



Drexel E-Repository and Archive (iDEA)

<http://idea.library.drexel.edu/>

Drexel University Libraries

[www.library.drexel.edu](http://www.library.drexel.edu)

The following item is made available as a courtesy to scholars by the author(s) and Drexel University Library and may contain materials and content, including computer code and tags, artwork, text, graphics, images, and illustrations (Material) which may be protected by copyright law. Unless otherwise noted, the Material is made available for non profit and educational purposes, such as research, teaching and private study. For these limited purposes, you may reproduce (print, download or make copies) the Material without prior permission. All copies must include any copyright notice originally included with the Material. **You must seek permission from the authors or copyright owners for all uses that are not allowed by fair use and other provisions of the U.S. Copyright Law.** The responsibility for making an independent legal assessment and securing any necessary permission rests with persons desiring to reproduce or use the Material.

Please direct questions to [archives@drexel.edu](mailto:archives@drexel.edu)

# Visualizing the Evolution of a Subject Domain: A Case Study

*Chaomei Chen*

Brunel University

*Leslie Carr*

Southampton University

## Abstract

We explore the potential of information visualization techniques in enhancing existing methodologies for domain analysis and modeling. In this case study, we particularly focus on visualizing the evolution of the hypertext field based on author co-citation patterns, including the use of a sliding-window scheme to generate a series of annual snapshots of the domain structure, and a factor-referenced color-coding scheme to highlight predominant specialties in the field.

**CR Categories and Subject Descriptors:** I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - Virtual Reality; H.3.7 [Digital Libraries]: Dissemination.

**Additional Keywords:** applications of visualization, domain analysis, citation analysis, visualization of literature

## 1 INTRODUCTION

More than 50 years ago, Vannevar Bush [1] envisaged his visionary device Memex. Like today's WWW, it would contain all sorts of information. Unlike the WWW, users would be able to build their own pathways or trails through the ever-growing information space — known as trailblazing. These trails would become part of the information space. This is a key concept concerning the accessibility and maintainability of a universal information space. In this case study, we focus on the role of information visualization techniques in capturing and representing underlying interrelationships in the literature of a subject domain.

With reference to the recent universal library initiative, Schatz [9] suggests that the next generation of information technology should transcend the boundary of documents and enable users to handle the semantics underlying these documents. Schatz and his collaborators analyzed bibliographic information obtained from several databases heavily used by computer scientists and created concept spaces on supercomputers.

1. Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3PH, UK. Tel: +44 1895 203080. E-mail: chaomei.chen@brunel.ac.uk
2. Department of Electronics and Computer Science, Southampton University, Southampton SO17 1BJ, UK. E-mail: L.Carr@ecs.soton.ac.uk

0-7803-5897-X/99/\$10.00 Copyright 1999 IEEE

The Institute for Scientific Information (ISI), best known for its Science Citation Index (SCI), has been exploring the structure of scientific literature based on citation data embedded in scientific literature for many years. ISI's work was originally motivated to break the barrier in subject indexing — by relying on the collective and accumulated views of researchers in a given discipline. Atlas of Science [7], the pioneering work at ISI, was generated according to document co-citation patterns. Recently, ISI is increasingly interested in the applicability of visualization technologies in revealing the structure of science [6, 11].

Author co-citation analysis (ACC) has been traditionally used to study the structure of a subject domain based on bibliographical data. ACC uses authors as data points in the literature. Its focus is on authors instead of articles or journals. More importantly, author co-citation is a more rigorous grouping principle than that of typical subject indexing, because it depends on repeated statements of connectedness by citers with subject expertise [13]. Author co-citations provide invaluable information about how authors, as domain experts, perceive the interconnectivity between published works, and what is a domain all about. A domain analyst can identify a sub-field based on the groupings of researchers who have contributed to closely related themes and topics. Such sub-fields are often known as specialties.

We originally introduced Pathfinder-based visualization techniques into author co-citation analysis in [5]. In this case study, we extend our approach further and particularly focus on visualizing the evolution of the hypertext field, including the use of a sliding-window scheme to generate a series of annual snapshots of the domain structure, and a factor-referenced color-coding scheme to highlight predominant specialties in the field.

## 2 RELATED WORK

Butterfly was designed as a user interface to science citation databases [8]. In Butterfly, the currently searched article is represented as the head of a butterfly. Citing and cited articles associated with the current article are represented on the wings of the butterfly. The design of Butterfly emphasized the role of an organic user interface — a virtual landscape grows under user control as information is accessed automatically.

White and McCain [13] used author co-citation analysis to map the field of information science. Their analysis included the top 120 authors ranked by citation counts drawn from 12 key journals in information science from 1972 through 1995. Their analysis clearly showed that the field of information science consists of two major sub-fields, experimental retrieval and citation analysis, and there was little overlap between their memberships.

Multidimensional scaling (MDS) techniques are commonly used in author co-citation analysis as a means of depicting the intrinsic interrelationships among a wide range of authors. However, MDS has a number of limitations as a network visualization solution as identified in [11]. Furthermore, due to the capacity of a general-

purpose statistical package, namely, SPSS, White and McCain had to limit the maximum number of authors to 100 authors in their MDS maps.

On the one hand, MDS has been widely used by domain analysts as well as other professionals. On the other hand, MDS also introduces additional complexities into domain analysis. For example, the nature of clustering is not always clear, and local details in an MDS map can be difficult to interpret.

In a series of studies, we have investigated the role of Pathfinder network scaling techniques in information visualization [2-5]. Pathfinder network scaling gives an analyst a greater control over extracting and representing the most salient structures defined by proximity data.

Author co-citation analysis provides an additional perspective to help us understand the dynamic structure of a field. We expect that author co-citation analysis, enhanced by visualization techniques, can play a significant part in helping people to make sense of a subject domain and the literature associated to the domain.

### 3 METHODS

In this case study, we incorporate Pathfinder-based visualization techniques into author co-citation analysis. In essence, the role of MDS is replaced and extended by using Pathfinder network scaling and virtual reality modeling techniques (see Figure 1).

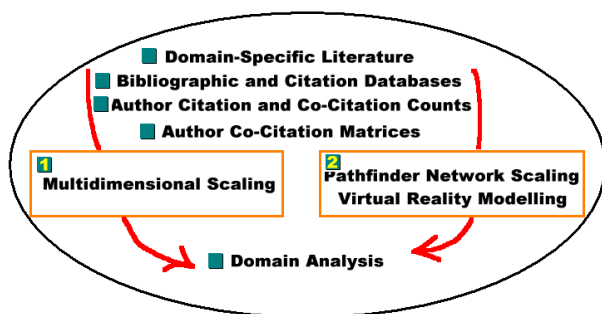


Figure 1: Methodologies of author co-citation analyses.

Author citation and co-citation counts were computed from a database containing all the articles published in the ACM Hypertext conference proceedings since 1989 up to 1998. Author co-citation counts were computed for all the authors who were cited five times or more during the whole period. This selection criterion resulted in a pool of 367 authors for the entire period.

In order to discover significant advances and trends in the history of the field, we applied the same methodology to a nine-year sequence of conference proceedings. We introduced a three-year sliding window scheme for these single-year visualizations. For year  $y$ , co-citation counts were calculated for authors who have been cited for five time or more in any year of the sliding window, i.e. year  $y-1$ ,  $y$ , or  $y+1$ . This sliding-window scheme provides a wider context for the author co-citation analysis in each single year.

Following [13], the raw co-citation counts were transformed into Pearson's correlation coefficients using the factor analysis on SPSS for Unix Release 6.1. These correlation coefficients were used to measure the proximity between authors' co-citation

profiles. Self-citation counts were replaced with the mean co-citation counts for the same author.

Pearson's  $r$  was used as a measure of similarity between author pairs, because, according to [13], it registers the likeness in shape of their co-citation count profiles over all other authors in the set. Pearson correlation matrices were submitted to Pathfinder network scaling. The resultant Pathfinder network was modeled in VRML. An author co-citation map was subsequently generated and incorporated both the author co-citation network and citation indices over three periods. Period I ranges from 1989 to 1991; Period II from 1992 to 1994; and Period III from 1996 to 1998. Hypertext reference links are provided in VRML versions of these maps. The name of each author in the co-citation map is linked to bibliographical details stored in a citation database accessible on the Web.

## 4 RESULTS

### 4.1 A Coherent View of Literature

Color plates 1 and 2 present two screenshots of the visualization of the hypertext subject domain. The visualization highlighted the most representative authors in the field and the strongest co-citation paths among them. Color plate 1 shows an overall author co-citation network derived from the entire time interval since 1989. The network is colored based on the results of a factor analysis of the co-citation matrix used in Pathfinder network scaling. Authors who have made generic contributions to the field tend to appear near to the central area of this network, whereas those who have made unique and specific contributions are likely to be found near the tip of a branch.

The skyline of a combined author citation and co-citation network echoes this interpretation (see Color Plate 2). In this combined network, periodical citations, as represented by stacked bars, are superimposed over the author co-citation map. An author's citations in each period are represented by a citation bar. The color of the bar indicates a specific period, and its height is proportional to the number of citations within this period. An interesting pattern is clear in this landscape — the closer an author is to the center, the more frequently this author has been cited.

### 4.2 The Evolution of the Hypertext Field

Tracking emergent trends and evolving patterns is one of the principal motivates for our visualization. Nine author co-citation maps are presented in Figure 2. Each one was generated automatically based on a three-year sliding-window model. Nodes at strategic positions in each map were manually annotated to highlight the essence of the citation patterns of the year. These strategic positions include nodes with high degree of connectivity, near to the center, or near to the tips. We labeled the names such as *Bush*, *Berners-Lee*, *Cailliau*, and *Rao* in the maps to trace their work. We also labeled some well-known systems in the maps to highlight significant contributions. The dynamics and evolution of the field can be analyzed visually. Authors' names are automatically available in the VRML version. These maps also provide direct access points to bibliographic and full text digital libraries. Researchers and analysts can use these maps directly in their citation analysis and domain analysis.

Figure 3 shows a 2D MDS map of top 100 most cited authors from the same data set. In Pathfinder networks, for example, the

network in Color Plate 1, branching nodes and explicit links provide valuable cues for analysts to interpret the network structure, whereas in the MDS map, such additional cues are not available.

### 4.3 Visualization and Hypertext

Information visualization is an inter-disciplinary domain. We are particularly interested in the nature of information visualization studies in the context of hypertext. The snapshot of 1998 author co-citation network (see Figure 4) reveals a number of clusters in association with information visualization, including names such as *Robertson*, *Mackinlay*, *Chalmers*, and *Fairchild*. Color-coded citation maps such as the one in Color Plate 1 tend to make the identification of such clusters easier.

The local details of one such cluster are included in Figure 4. To some extent, the members of this cluster reflect the nature of the impact of their work to the hypertext field. For example, *Chalmers* was often cited for his work in BEAD, *Kamada* and *Kawai* for their graph layout algorithm, *Fairchild* and *Poltrock* for

SemNet, and *Zizi* for her work in SHADOCS. Both BEAD and SemNet were published outside the hypertext literature. The names of *Salton*, *Buckley*, and *Singhal* imply that these visualization works probably have a profound connection to the famous vector space information retrieval model. In deed, the classic vector space model and its various variations have a substantial impact on the advances of information visualization.

## 5 DISCUSSIONS

In this case study, we have generated a number of author co-citation maps of the field of hypertext. These maps not only visualize the structure and evolution of the hypertext literature, but also provide a gateway for users to access the literature of hypertext directly. These overview maps currently provide hyperlinks from an author to corresponding entries of the author in a bibliographic database accessible on the WWW. These hyperlinks can be automatically reconfigured so that from these maps one can directly access full-text articles in the ACM Digital Library on the WWW.

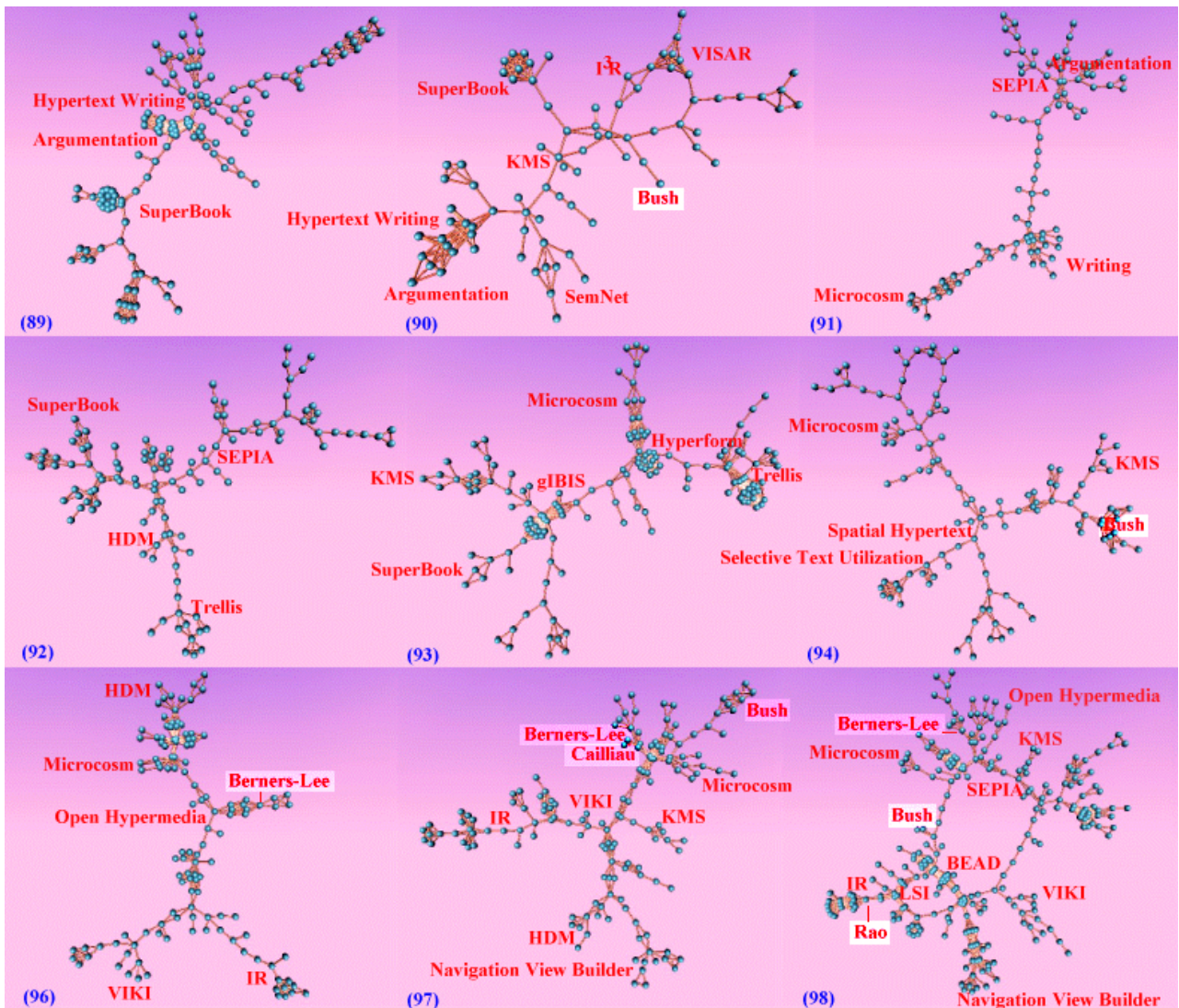


Figure 2: Nine author co-citation networks of the hypertext field (1989-1998).

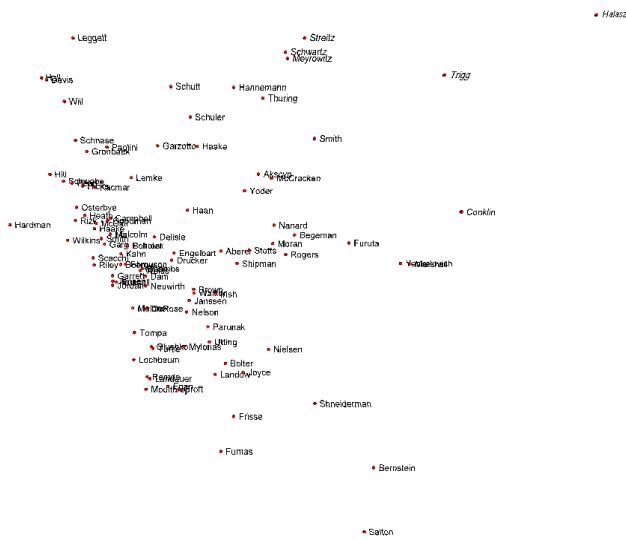


Figure 3: Top 100 most cited authors in a 2D MDS map.

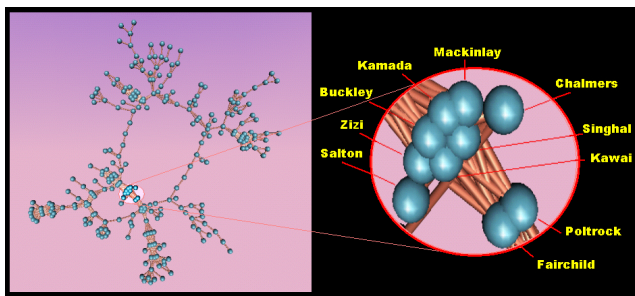


Figure 4: A cluster of authors associated with visualization in the 1998 author co-citation map.

This case study is based on the ACM Hypertext conference proceedings alone. Of course, there are many other forums for hypertext research and development. In general, interdisciplinary subject domains present opportunities as well as challenges. We plan to extend the scope of our domain mapping to incorporate information from several disciplinary-specific sources. In particular, we are interested in mapping the field of information visualization based on the IEEE visualization conferences, the WWW conferences, the ACM SIGGRAPH and SIGIR conferences, and key journals in these fields. Interdisciplinary literature mapping is likely to reveal more insightful patterns than a single-domain literature mapping. More importantly, it provides a useful tool for researchers and analysts to explore the literature of a subject domain and extend our knowledge from one field to another.

## 6 CONCLUSION

In this case study, we have integrated visualization techniques into author co-citation analyses. We have used Pathfinder networks to layout our maps. These maps are different from the MDS-based maps typically used in author co-citation analyses, such as [13]. Pathfinder networks can provide more accurate information about local structures than MDS maps [10]. We found that the provision

of explicit links in our maps made it easier to interpret interrelationships among different data points.

The value of this work is in its ability to thread through the literature and extract the most salient associations among authors who have made significant contributions to the field. Furthermore, author co-citation maps provide a means of identifying research fronts, i.e. specialties in the field, and a visual aid of interpreting the results of citation analysis.

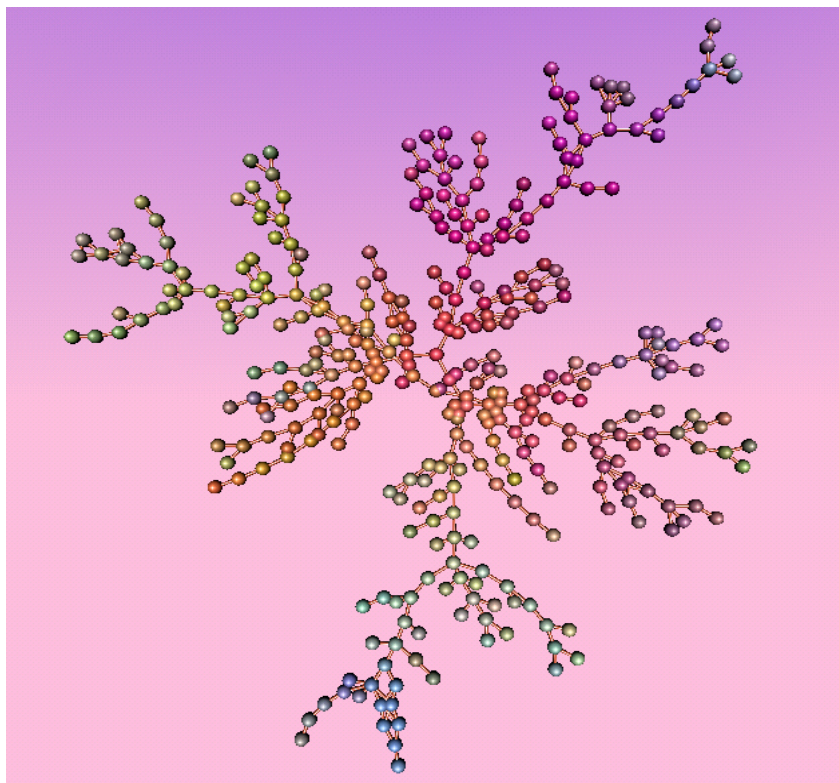
This case study has provided us invaluable experience in synthesizing the literature of the hypertext field and extending domain modeling and analysis methodologies with information visualization techniques.

## Acknowledgements

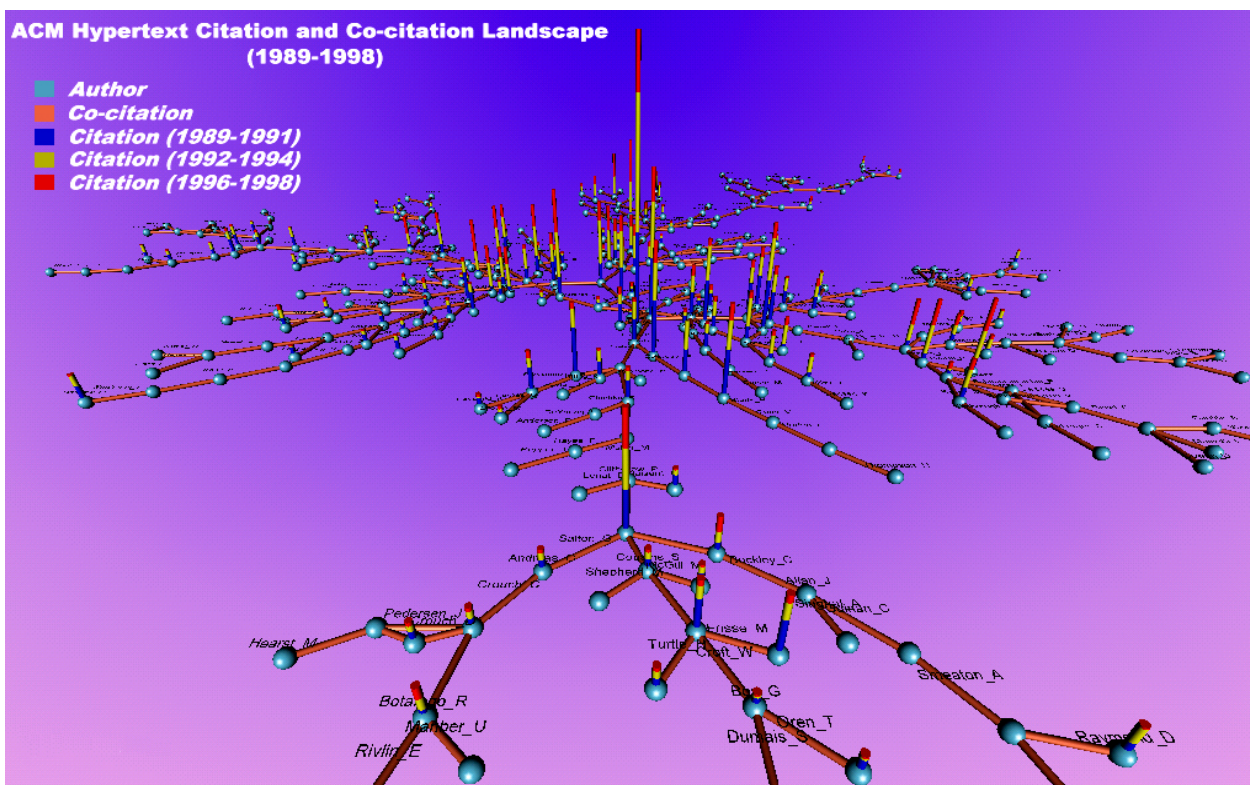
This work was in part supported by the British research council EPSRC under the Multimedia and Networking Applications Programme (Research Grant: GR/L61088).

## References

- [1] Vannevar Bush. As we may think. *The Atlantic Monthly*, 176(1), 101-108, 1945.
- [2] Chaomei Chen. Bridging the gap: The use of Pathfinder networks in visual navigation. *Journal of Visual Languages and Computing*, 9(3), 267-286, 1998.
- [3] Chaomei Chen. Generalised Similarity Analysis and Pathfinder Network Scaling. *Interacting with Computers*, 10(2), 107-128, April 1998.
- [4] Chaomei Chen. *Information Visualisation and Virtual Environments*. Springer-Verlag, 1999. ISBN 1-85233-136-4.
- [5] Chaomei Chen and Leslie Carr. Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998). In *Proceedings of the 10th ACM Conference on Hypertext and Hypermedia*, pages 51-60. February 1999.
- [6] Eugene Garfield. Mapping the world of science. In *Proceedings of the 150 Anniversary Meeting of the AAAS*, Philadelphia, PA, pages 1-19, February 1998.
- [7] Institute for Scientific Information. *ISI atlas of science: Biochemistry and molecular biology, 1978/80*, Institute for Scientific Information, Philadelphia, PA, 1981.
- [8] Jock D. Mackinlay, Ramana Rao, and Stuart K. Card. An organic user interface for searching citation links. In *Proceedings of ACM Conference on Human Factors in Computing Systems*, pages 67-73, 1995.
- [9] Bruce R. Schatz. Information retrieval in digital libraries: Bringing search to the net. *Science*, 275, 327-334, 1997.
- [10] Roger W. Schvaneveldt, F. T. Durso, and D. W. Dearholt. Network structures in proximity data. In G. Bower, editor, *The Psychology of Learning and Motivation*, 24, Academic Press, 249-284, 1989.
- [11] Henry Small. Update on science mapping: Creating large document spaces. *Scientometrics*, 38(2), 275-293, 1997.
- [12] Stephen G. Eick and Graham J. Wills. Navigating large networks with hierarchies. In *Proceedings of IEEE Visualization '93 Conference*, pages 204-210, 1993.
- [13] Howard D. White, and Katherine W. McCain. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49(4), 327-356, 1998.



Color Plate 1: An author co-citation network (ACM Hypertext 1989-1998). The network is colored by each author's membership of three predominant sub-fields.



Color Plate 2: A combined citation and co-citation landscape of the field. Citations over three consecutive periods (colored stacked bars) are superimposed over the co-citation network of the entire history.