Six Degrees of Jonathan Grudin: A Social Network Analysis of the Evolution and Impact of CSCW Research

Daniel B. Horn, Thomas A. Finholt, Jeremy P. Birnholtz, Dheeraj Motwani, and Swapnaa Jayaraman
School of Information, University of Michigan
1075 Beal Avenue
Ann Arbor, MI 48109
+1 734 764 6131

{danhorn, finholt, jbirnhol, dmotwani, swapnaa} @umich.edu

ABSTRACT

In this paper, we describe the evolution and impact of computer-supported cooperative work (CSCW) research through social network analysis of coauthorship data. A network of authors as nodes and shared papers as links is used to compare patterns of growth and collaboration in CSCW with other domains, such as high-energy physics and computer science. Further, the coauthorship network data are used to depict dynamic changes in the structure of CSCW collaborations over time. Examination of these changes shows high volatility in the composition of the CSCW research community over decade-long time spans. These data are augmented by a brief citation analysis of recent CSCW conferences. We discuss the implications of the CSCW findings in terms of the influence of CSCW research on the larger field of HCI research as well as the general utility of social network analysis for understanding patterns of collaboration.

Categories and Subject Descriptors

K.2 [**Computing Milieux – History of Computing**]: – *people, theory*. J.4 [**Computer Applications – Social and Behavioral Sciences**] – *sociology, psychology*. H.5.3 [**Information Systems -- Information Interfaces and Presentation**]: Group and organization interfaces – *computer-supported cooperative work, evaluation/methodology*.

General Terms

Measurement, Theory.

Keywords

Social network analysis, coauthorship, collaboration, computer-supported cooperative work

1. INTRODUCTION

The field of computer-supported cooperative work (CSCW) has an intense interest in studying collaborative practices, yet ironically, CSCW researchers remain unreflective about the structure and impact of their own collaborations. This indifference is in contrast to recent efforts in other disciplines, notably physics, where there is a growing literature on the organization and evolution of collaborations [4, 25]. Social network analysis is the primary lens used to understand patterns of collaborations in these other fields.

Sociologists have understood for some time that social relations within a group, such as friendship ties, can be represented as graphs (e.g. [28]). In these graphs, nodes represent group members and links represent connections between the members – such as frequency of communication, workflow exchanges and so forth. In the 1960s and 1970s, researchers developed computer programs to analyze characteristics of groups based on network properties. For example, members of a network differ in terms of their centrality in a network. An individual with high centrality is potentially influential because this person may link together many people who otherwise wouldn't be connected. Recently, advances in the capabilities of network analysis software combined with increased computational power have allowed examination of very large networks, sometimes with tens or hundreds of thousands of nodes or more. These techniques have been used to understand a wide variety of systems including networks of high-school friendships [22] Internet hosts [8] and the interactions of molecules in a cell [17].

The ability to examine very large networks means that it is now possible to view ecologies of collaborations – such as entire scientific disciplines. This enterprise is well underway in disciplines like physics, where Newman and others have used a variety of metrics to compare features of collaboration across disciplines [4, 24, 25]. For instance, one measure is the relative size of the *giant component*, which captures the proportion of all researchers in a given domain who are connected to the largest common network. One interpretation of the size of the giant component is that it provides a relative index of the degree of integration within a field. Other popular measures relate to the geodesic properties of a network, well-known from Milgram's famous "small world" studies documenting that any two randomly selected individuals have a median separation of six people [21]. Dodds, Muhamed, and Watts [10] have confirmed this result for email associations – and of course the concept has been popularized through the play and film *Six Degrees of Separation*, the popular Kevin Bacon game (i.e., most movies stars can be linked through co-stars to Kevin Bacon in a small number of steps, [26]), and mathematicians' calculation of their Erdös numbers (i.e., their proximity, in terms of coauthorship, with the famous itinerant mathematician, Paul Erdös, [13]). One indication of just how small the world of academic researchers is can be seen in the fact that the authors of the present paper have Erdös numbers which range from five to six – even though none are even remotely connected with mathematics.

In addition to comparisons across fields, techniques for analyzing very large networks are also useful for comparisons across time within a field. For example, by tracking network properties such as the relative magnitude of the giant component, it is possible to measure the rate and extent of consolidation over time. In particular, there are two ways of viewing these dynamic network properties. Barabasi et al. [4] examine the accumulation of interconnections over time by generating a series of cumulative networks, one for each year that includes the current and all previous years (i.e., one network for 1991-1992, the next for 1991-1993 and so on). This assumes that all links have the same value, such that a dormant collaboration is as significant as an active collaboration. Another view, which we develop in this paper, is that the level and structure of interconnections change over time, and therefore the dynamic network should be viewed as a sliding snapshot of a field. This approach involves generating a series of networks that each represent the field at a specific point in time (e.g., examining all collaborations reflected in the publication record for a particular year). However, one challenge that this presents is that the network has no "memory" of recent collaborations – a pair of researchers who collaborate only in years 1 and 3 would not appear to be collaborators in year 2. To overcome this problem, we assume that a collaborative link persists for a time, and that if it is not renewed through additional collaborative productivity, it decays. Through an informal examination of HCI coauthorship records, we have settled on a five-year period of collaboration persistence. One way to think of this is as a step decay function. That is, when two or more people collaborate on a paper, they become linked, and because these collaborative relationships may change over time, we remove the link between two people if more than five years elapse without the pair producing another collaborative work. This can also be thought of as viewing the network over a series of overlapping five-year windows, within which, the network represents the pattern and interconnection of active researchers during that period. One advantage of this perspective is that it provides a clearer picture of the succession of different generations of researchers and of the movement by researchers in and out of a given field.

Given the tools and measures described above, the interesting question becomes how to use these techniques to address hypotheses about the formation, structure, and impact of the CSCW research community. The motivation for examining CSCW researchers is twofold. First, as members of this community, we are curious about the origins and elaboration of the CSCW field. Second, in a more general sense, the emergence of CSCW research is an instance of the broader phenomenon of new disciplines forming at the intersection of existing fields. In particular, numerous recent accounts of the organization of scientific activity highlight the importance of cross- and multi-disciplinary research – with a number of calls to make multi-disciplinary research the norm [11].

1.1 Origins and elaboration of CSCW

CSCW emerged as a field in response to the recognition that group and organizational contexts matter in human-computer interaction (HCI). Earlier attempts to incorporate group and organizational themes, such as the office automation movement in the 1970s and 1980s, failed to reach critical mass. However, researchers began to coalesce more successfully starting with a 1984 workshop, where Irene Greif coined the term "computer-supported cooperative work." The 1984 workshop led to the first CSCW conference in 1986 – and from there the further crystallization of a research community that had previously existed within the larger human-computer interaction community and within other disciplinary communities, such as anthropology, computer science, and psychology [3]. Early participants in the CSCW field formulated and followed a research philosophy that stressed the benefit of system development guided by insights about the social context of both development and use. During this formative period, there was openness to multiple theoretical and methodological approaches [15]. For example, early CSCW conferences drew on a wide variety of sources, as measured by citation diversity (e.g. [12]). Over time, however, this openness declined as certain theoretical and methodological perspectives became more dominant, such as the prominence of ethnomethodology. In addition, the field continues to struggle with core tensions – such as how to weight technical versus social contributions, and how to maintain relevance to broader communities while building a field identity [14].

In this paper we examine issues in the birth and development of CSCW through social network analysis of coauthorship networks. These analyses were motivated by three specific aims.

First specfic aim. We wanted to examine the extent to which CSCW researchers have maintained ties within other fields over time. Specifically, if CSCW is a broad home to many different disciplines this should be reflected in the breadth of ties by researchers within CSCW. For example, during the founding era of the CSCW conferences, authors came from many home disciplines – and this is often celebrated as a virtue of the of CSCW approach. However, as sub-fields mature, they can also draw firmer boundaries – sometimes at the expense of closing potentially useful connections to interesting related fields. Through analysis of the proportion of each CSCW researcher's coauthors who were CSCW researchers themselves – we were able to create a picture of how cosmopolitan CSCW research was at any given moment between 1986 and 2003.

Second specific aim. We wanted to examine the stability of the CSCW community composition over time. Specifically, if engaging new ideas periodically refresh CSCW research, this should be reflected in a corresponding influx of new researchers. For example, the consensus around ethnomethodology brought in a large number of European researchers in the late 1980s – and led to the founding of a parallel European CSCW conference in 1989. Too much churn in composition, however, can undermine formation of field identity – as researchers come and go too rapidly to make community-building contributions. Similarly, lack of turnover can lead to stasis and irrelevance, as the field fails to recruit new participants and remains preoccupied with old problems. Through analysis of the year-to-year appearance of new CSCW researchers – we were able to create a picture of the stability of the CSCW community over the period 1986 to 2003.

Third specific aim. Finally, we wanted to examine the visibility of the CSCW community over time. Specifically, sub-fields emerge because they want to pursue agendas that differ from the larger home fields. Yet, the sub-fields also want to continue to influence the larger fields. For example, one reading of the formation of the CSCW community is that a group of participants in the HCI community wanted to have a more explicit focus on group and organizational aspects of human-computer interaction. However, to the extent that some participants in CSCW wish to change broad practices, such as the design of human-computer interfaces in general – not just for group applications – then CSCW research must remain visible and vital to the larger HCI community. Through analysis of the centrality of CSCW researchers within the HCI community we were able to gauge the influence of CSCW researchers within HCI, such as the rank of central CSCW researchers among all HCI researchers.

2. METHOD

2.1 Data

The data for this study came from the database of HCI publications supported by ACM and maintained by Gary Perlman at http://www.hcibib.org, which includes entries for journal articles, books, book chapters, conference proceedings, videos, and web sites [27]. With Gary's permission, we downloaded records for the period 1982 through the present (as of January 30, 2004). We chose 1982 as the start date since this coincides with the first SIGCHI conference – marked by most as the formal beginning of the HCI field. For each record, we were interested in year of publication, publication outlet (i.e., the conference, book or journal), and the article authors. For our purposes, we included only journal articles, conference proceedings, books, and book chapters in our analyses. We analyzed a total of 22,887 publications.

A significant complication of using the bibliographic data is determining unique name identifiers for authors. That is, "Judith S. Olson," "Judy Olson," and "J.S. Olson" all refer to the same person – yet these multiple identities must be resolved to a single identifier for purposes of assigning network nodes (otherwise each identity will be treated as a separate person). In addition to variations in the way one's name is reported, typos and Optical Character Recognition (OCR) errors, as well as name changes occur as well. For example, "Judith S. Olsen" and "Judith Reitman Olson" are also the same person. The problem of name matching is one faced in a variety of contexts, and one which is surprisingly challenging for computer algorithms [6, 16]. Our sample contained 27,408 unique strings representing author names. Through a combined process of algorithmic and manual techniques, we allocated those strings to 23,624 individuals. In a small number of cases, such as for undifferentiated common names (e.g., "J. Smith" may be "John Smith" or "Jane Smith"), we were unable to disambiguate the names. One beneficial side effect of this process was the identification of 185 name errors in the database (as well as 6 misspelled author names in the papers themselves!) – a remarkably small error rate given the number of entries. These errors were corrected in our analyses, and forwarded to hcibib in order to update their database.

2.2 Analysis

The disambiguated records from hcibib were used to make lists of pairwise connections for input to network analysis packages. For example, an article with three authors (A, B, C) would result in a list of three dyadic entries (i.e., AB, AC, BC). Typically, a network analysis package, such as *Pajek* [5], takes a pairwise list and builds an edge between each pair, such that the entry "AB" would be interpreted as "build a link between node A and node B." Across all such links we can compute summary network statistics as well as draw two- or three-dimensional visualizations of the resulting network. In this paper we focused on a small set of network measures. First, we are interested in statistics that characterized an entire network – such as the diameter (described earlier in terms of the small world phenomenon). Second, we were interested in statistics that characterized nodes within a network – such as centrality. Centrality is often described in multiple ways. Here we examined "degree centrality," or simply the total number of links at any node, and "betweeness centrality," or the number of shortest paths between any two other nodes that pass through a target node. Roughly speaking, centrality defines a kind of influence or visibility in the network. For both network and node measures we examined changes over time by applying a sliding five-year window. Finally, for analyses specific to the CSCW community, we defined a CSCW researcher as anyone who had published in a CSCW-related outlet (i.e., *Proceedings of CSCW*, *Proceedings of ECSCW*, or the journal *Computer-Supported Cooperative Work*). The CSCW giant component was defined as the largest connected network, within any five-year window, of people defined as CSCW researchers. Note that the largest component could include two people defined as CSCW researchers, who were coauthors on a publication in a non-CSCW outlet (e.g., CHI or UIST). In any five-year window, a person was considered to be a new CSCW researcher if they had not previously appeared as a member of the CSCW giant component.

3. RESULTS

3.1 HCI compared to other domains

To understand how CSCW coauthorship matched up against other disciplines documented in the literature, we conducted a number of comparative analyses. For these comparisons, we were restricted to the larger HCI community – since the rate and number of publications within the smaller CSCW community is not sufficient to produce meaningful comparisons. Table 1 summarizes the comparisons we performed. Following the work of Newman [25], we compared HCI coauthorship patterns to patterns in high-energy physics (HEP), computer science (CS), and biomedical research. The columns in Table 1 show the different scientific communities. The rows correspond to measures used to compare the communities. There is a clear qualitative difference between the HEP and biomedical communities – and the CS and HCI communities. Specifically, in HEP and biomedical research, papers have a higher average number of coauthors – and the relative size of the giant component in these fields is larger. As the table shows, within HCI, 51.3% of all authors are connected through an average of 6.8 steps (in reference to this paper's title, 94.7% of those in the HCI giant component have Jonathan Grudin numbers of 6 or less).

3.2 CSCW compared to HCI

3.2.1 Links to the broader HCI community

Consistent with our first aim, we examined the extent to which the CSCW community has maintained links to the larger HCI community over time.

Table 1. Comparison of HCI community characteristics to other disciplines

	Field			
Measure	**Biomed[a]**	**HEP[a]**	**CS[a]**	**HCI**
Number of authors	1,520,251	56,627	11,994	23,624
Number of papers	2,163,923	66,652	13,169	22,887
Papers per author	6.4	11.6	2.6	2.2
Authors per paper	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
Largest distance	24	19	31	27

Note: [a] data from Newman [25]

To conduct this analysis we identified CSCW researchers and then examined the composition of links for each person. Links were classified as "CSCW-related" if they went to another CSCW researcher, and as "non-CSCW" if they went outside the CSCW network. Within each sliding five-year window, starting in 1986, we computed the mean proportion of CSCW links for each CSCW researcher. The results of this analysis are shown in Figure 1. Note that the trend over time is steady. That is, CSCW researchers had a high proportion of coauthors outside the CSCW community when the field emerged in the late 1980s, and they continue to maintain a large number of connections to coauthors in other fields. Of the 48 people in the CSCW giant component for the period 1986-90, the average person had 3.9 CSCW coauthors and 3.7 non-CSCW coauthors. During the most recent five-year period, 1999-2003, of the 188 people in the CSCW giant component, the average person had 5.1 CSCW coauthors and 4.5 non-CSCW coauthors.

3.2.2 *Author rankings in CSCW and HCI*

These coauthorship networks allow us to conduct analyses at various levels. We can describe the field as a whole, as we do in Table 1, and we can do more fine-grained analyses at the level of communities or individuals. There are a number of metrics that can be calculated for individuals in the network. In Table 2, we present the top 20 individuals or pairs in CSCW and HCI on measures of centrality, productivity, and collaboration strength for the cumulative network from 1982 to 2003

There are numerous centrality measures that can be calculated for an individual in a network, we have used two measures. The first, *degree centrality* ("Number of Coauthors" in Table 2) indicates the number of nodes to which a particular node is directly linked. In our networks, degree centrality represents the number of collaborators an individual has had. An individual with high degree centrality is central in the sense that he or she is directly connected to a large number of other people. A second measure, *betweenness centrality* reflects the number of shortest paths between pairs of nodes that pass through a particular node. A person with high betweenness centrality is central in the sense that he or she may control the flow of information between different communities. In Table 2, the betweenness scores are calculated from the giant components in the HCI network and CSCW network respectively. Productivity is represented in Table 2 by the number of papers published by an individual.

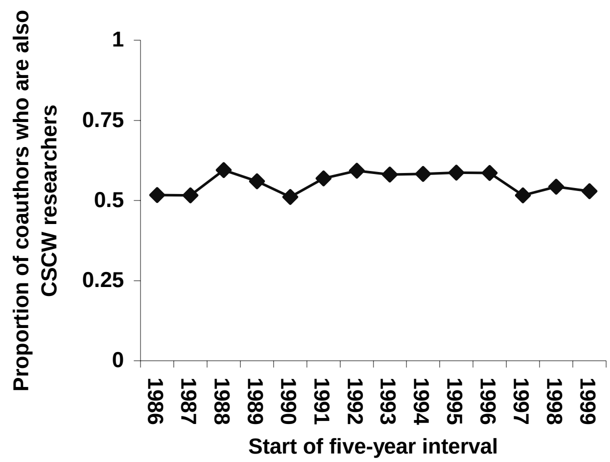

Figure 1. Variation over time in the mean percentage of CSCW researchers in the coauthorship networks of people in the CSCW giant component.

In order to examine the strength of collaborations in our networks, we use a collaboration weighting metric proposed by Newman [25], wherein the strength of the collaborative link between two authors on a paper is inversely related to the number of total authors on the paper. That is, for a paper with n authors, the collaboration weight between each pair of authors increases by $1/(n\text{-}1)$. For each paper that two individuals coauthor, their collaborative strength increases relative to the number of other individuals in the collaboration. By adding these collaborative weights across all papers a pair produces, we have a measure of the collaborative strength of the pair. Table 2 shows the top pairs in terms of collaborative strength. A pair with a high collaborative strength has written numerous papers together, with relatively few coauthors.

In addition to looking for one's own name in Table 2, one can use these lists to examine patterns within and between CSCW and HCI. Among the observations that we have made:

Proximity and Collaboration. Only four of the top 20 (indeed, of the top 50) collaborative pairs in HCI have only published in the HCI literature while at different institutions. This is in line with others' findings on the role of physical proximity and collaboration [19], and interesting given that one would expect that our community would be more likely than most to be familiar with collaborative technologies. Additionally, in contrast to what would be expected in many other fields at least four of the top 30 collaborative pairs are married couples (to the best of our knowledge).

Cohort Effects. Inspection of Table 2 makes it clear that most of the top ranked individuals have been active in the HCI and CSCW communities since the 1980s. Obviously, a longer tenure affords more opportunities for production and collaboration, and may make it difficult to compare the impact of these veterans with the field's relative newcomers. In order to level the playing field, we

Table 2 – The authors with the highest numbers of papers, numbers of coauthors, and betweenness centrality scores, and the strongest collaborations in HCI and CSCW. These data are based on the cumulative database from 1982 to 2003. Full lists of the rankings of all the others in these databases can be found on the world-wide web [30]. (Note that betweenness centrality scores have been multiplied by 100)

		Number of Papers		Number of Coauthors		Betweenness (X 100)		Collaboration Strength	
HCI	1	Ben Shneiderman	127	Ben Shneiderman	167	Ben Shneiderman	10.25	M. B. Rosson/J. M. Carroll	20.7
	2	John Carroll	113	Jakob Nielsen	104	William Buxton	6.19	C. Stephanidis/D. Akoumianakis	16.5
	3	Steven Pemberton	108	John Carroll	102	Gavriel Salvendy	5.71	K. A. Holtzblatt/H. R. Beyer	16.1
	4	Gavriel Salvendy	98	Allison Druin	102	James Foley	5.65	C. Stephanidis/A. Savidis	13.8
	5	Jakob Nielsen	92	Elliot Soloway	93	Jakob Nielsen	5.60	J. C. Scholtz/S. Wiedenbeck	12.0
	6	Brad A. Myers	84	Michael Muller	92	John Carroll	4.57	J. Grudin/S. E. Poltrock	11.9
	7	Jonathan Grudin	66	Brad Myers	89	Thomas Landauer	3.57	S. Greenberg/M. Roseman	11.5
	8	Saul Greenberg	64	Gavriel Salvendy	81	Brad Myers	3.48	G. M. Olson/J. S. Olson	11.3
	9	William Buxton	62	William Buxton	79	Thomas Moran	3.40	M. E. Atwood/G. A. Boy	11.0
	10	C. Stephanidis	62	Steven Benford	74	Jonathan Grudin	3.09	B. Shneiderman/C. Plaisant	10.3
	11	Gary Perlman	59	Elizabeth Mynatt	72	Austin Henderson	2.92	B. R. Gaines/M. L. G. Shaw	10.0
	12	Alistair Sutcliffe	58	Terry Winograd	72	Marilyn Mantei	2.80	R. K. Furuta/P. D. Stotts	9.5
	13	Mary Beth Rosson	57	Hiroshi Ishii	71	Hiroshi Ishii	2.80	H. R. Hartson/D. Hix	9.4
	14	Andrew Monk	57	Catherine Plaisant	71	Susan Dray	2.70	C. Heath/P. Luff	9.3
	15	Marisa Campbell	57	Jonathan Grudin	70	Terry Winograd	2.60	S. Greenberg/C. Gutwin	9.2
	16	Stuart Card	55	Dan Olsen	69	Gerhard Fischer	2.56	L. Candy/E. A. Edmonds	9.0
	17	Bonnie John	53	John Karat	69	Jean Scholtz	2.46	P. Johnson/H. Johnson	8.8
	18	Allison Druin	53	James Landay	65	Jennifer Preece	2.28	G. P. Kurtenbach/W. A. S. Buxton	8.5
	19	Michael Muller	52	Claire O'Malley	65	Gary Marchionini	2.27	P. Palanque/R. Bastide	8.3
	20	Thomas Moran	51	James Foley	64	John Karat	2.18	S. Benford/C. Greenhalgh	8.3
		Elizabeth Mynatt	51						
CSCW	1	John Carroll	113	Jakob Nielsen	104	Jonathan Grudin	10.90	J. Grudin/S. E. Poltrock	11.9
	2	Jakob Nielsen	92	John Carroll	102	Tom Rodden	10.38	S. Greenberg/M. Roseman	11.5
	3	Brad Myers	84	Michael Muller	92	Paul Dourish	9.43	G. M. Olson/J. S. Olson	11.3
	4	Jonathan Grudin	66	Brad Myers	89	Thomas Moran	8.07	R. K. Furuta/P. D. Stotts	9.5
	5	Saul Greenberg	64	Steven Benford	74	Bonnie Nardi	8.00	C. Heath/P. Luff	9.3
	6	Michael Muller	52	Elizabeth Mynatt	72	Robert Kraut	7.80	S. Greenberg/C. Gutwin	9.2
	7	Thomas Moran	51	Terry Winograd	72	Lucy Suchman	7.15	S. Benford/C. Greenhalgh	8.3
	8	Elizabeth Mynatt	51	Hiroshi Ishii	71	Sara Bly	6.69	D. McCracken/R. M. Akscyn	8.0
	9	Dan Olsen	50	Jonathan Grudin	70	Terry Winograd	6.67	L. G. Terveen/W. Hill	8.0
	10	Scott Hudson	50	Dan Olsen	69	Judith Olson	6.65	A. Girgensohn/A. Lee	7.6
	11	Hiroshi Ishii	50	John Karat	69	Liam Bannon	6.52	D. A. Henderson, Jr./K. Ehrlich	6.6
	12	John Karat	48	Claire O'Malley	65	Christian Heath	6.49	J. M. Haake/W. Wang	6.3
	13	Tom Rodden	48	Tom Rodden	64	Marilyn Mantei	6.46	W. K. Edwards/E. D. Mynatt	6.2
	14	Austin Henderson	45	Robert Kraut	63	John Carroll	6.22	I. Smith/S. E. Hudson	5.7
	15	Jean Scholtz	43	Judith Olson	60	Irene Greif	6.09	F. M. Shipman, III/C. C. Marshall	5.6
	16	Robert Kraut	42	Scott Hudson	59	Gloria Mark	5.06	R. Choudhary/P. Dewan	5.5
	17	Steven Benford	42	Marilyn Mantei	58	Steve Whittaker	4.93	F. Ljungberg/S. Kristoffersen	4.8
	18	Alan Dix	40	Thomas Moran	56	Thomas W. Malone	4.87	J. M. Carroll/M. K. Singley	4.7
	19	Gary Olson	40	Norbert Streitz	56	Scott Hudson	4.42	S. R. Hiltz/M. Turoff	4.5
	20	Allan Maclean	39	Austin Henderson	55	Gary Olson	4.07	R. E. Kraut/R. S. Fish	4.5

have divided the CSCW community into three major cohorts based on when each individual first appears as part of the CSCW giant component. The first cohort includes individuals who first appeared between 1988 and 1992. Members of the second cohort first appeared between 1993 and 1997. The third cohort consists of individuals who first appeared in the CSCW giant component between 1998 and 2002. Tables 3 and 4 show the ten individuals from each cohort with the greatest betweenness scores for each non-overlapping five year "generation." It is important to note that cohort inclusion is based on the year an individual first entered the giant component in the CSCW community. This means that an individual may have published papers (either in CSCW outlets or not) earlier than his or her cohort assignment suggests, but was not connected via coauthorship to the giant component until later.

Table 3. Members of the 1988-1992 CSCW cohort with the highest betweenness centrality scores over time.

	Betweenness Scores for 1988-1992 Cohort		
	1988-1992	1993-1997	1998-2002
1	Robert Kraut	Terry Winograd	Robert Kraut
2	Thomas Malone	John Patterson	Bonnie Nardi
3	Gary Olson	Robert Kraut	Judith Olson
4	Pelle Ehn	Jonathan Grudin	Michael Muller
5	Jonathan Grudin	Sara Kiesler	Jonathan Grudin
6	Irene Greif	Liam Bannon	Hiroshi Ishii
7	Christine Neuwirth	Mike Robinson	Christine Neuwirth
8	Tora Bikson	Kjeld Schmidt	James Morris
9	Michael Muller	Christine Neuwirth	Lucy Suchman
10	Lucy Suchman	Hiroshi Ishii	John Tang

3.2.3 Stability of the CSCW community

Consistent with our second aim, we examined the stability of the CSCW community over time. To conduct this analysis we identified members of the CSCW community, as defined above, and analyzed the size and composition of the CSCW giant component in sliding five-year windows anchored on 1984. The results of this analysis are shown in Figure 2. Note that the giant component grew steadily through the five-year period 1996-2000, when there were 294 CSCW researchers in the largest connected network. From this peak, however, the size of the giant component dropped by 36% to 188 researchers for the five-year period 1999-2003. The decline in the size of the giant component occurred despite a relatively steady influx of new researchers into the giant component, i.e., within the range of 60 to 120 new people in each interval. In addition, the HCI giant component grew by 16% over the same interval, from 3033 researchers in the five-year period 1996-2000, to 3517 researchers in the interval 1999-2003.

3.2.4 Visibility of the CSCW community

Consistent with our third aim, we examined the visibility of the CSCW community. To conduct this analysis we compared the betweeness centrality of members of the CSCW community with their corresponding betweeness centrality in the larger HCI community. We were interested in knowing whether people highly visible (i.e., highly central) in the CSCW community were also central in the HCI community. Figure 3 shows the results of this analysis. Note that in general, highly central people in CSCW were also highly central in the HCI community. However, there is a sub-set of CSCW researchers (below the diagonal) who were central to CSCW, but much less central to HCI. Similarly, there is a sub-set of CSCW researchers less central to CSCW, but very central to HCI.

Table 4. Members of the 1993-1997 and 1998-2002 CSCW cohorts with the highest betweenness scores over time.

	Betweenness Scores for		
	1993-1997 Cohort		1998-2002 Cohort
	1993-1997	1998-2002	1998-2002
1	Paul Dourish	Gloria Mark	Erin Bradner
2	Sara Bly	Paul Dourish	Thomas Erickson
3	Steve Whittaker	Keith Edwards	Rebecca Grinter
4	Wolfgang Prinz	Steve Whittaker	Elizabeth Mynatt
5	Giorgio De Michelis	Tom Rodden	Joseph Konstan
6	Jakov Kucan	Thomas Moran	JJ Cadiz
7	Tom Brinck	Christian Heath	Judith Donath
8	Susan Mcdaniel	Vicky O'Day	Darren Gergle
9	John Bowers	Ellen Isaacs	Shahram Izadi
10	Gloria Mark	Jane Siegel	Steven Poltrock

3.3 Diffusion of ideas

Centrality as measured through coauthorship networks may reflect one kind of visibility and influence, but there may be other forms of centrality that are equally or more important. For example, citation impact is a typical measure of influence. That is, if papers formed the nodes and citations formed the links, papers with high centrality would be those that are frequently referenced by other papers. This kind of centrality is often weighted more heavily in terms of reputation and promotion than coauthorship centrality. However, review of the top twenty most central people in CSCW suggests that there is a likely strong relationship between citation centrality and coauthorship centrality – since the list of those with high coauthorship centrality would certainly overlap with typical lists of the highest impact researchers in the field.

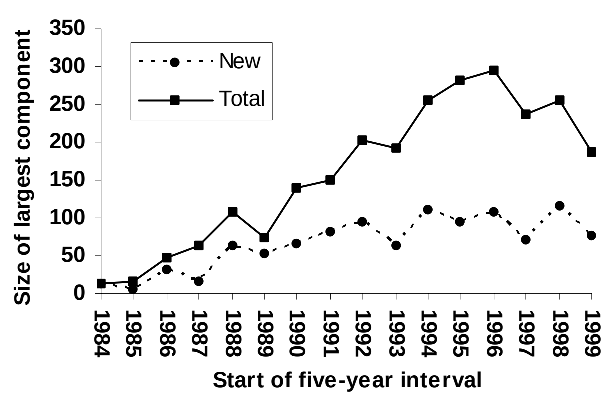

Figure 2. Change over time in the number of members and number of new members in the CSCW giant component

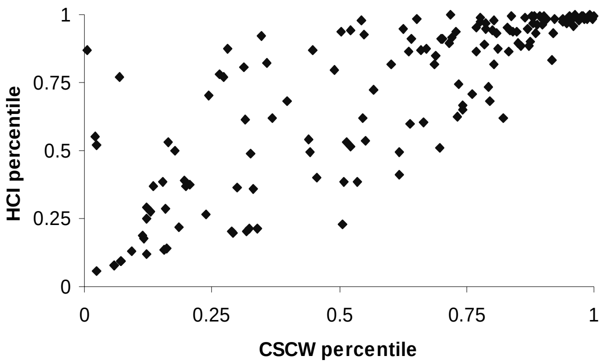

Figure 3. Plot of CSCW centrality percentile rank against HCI centrality percentile rank for the HCI and CSCW cumulative coauthorship networks, 1982-2003

Although a full-scale citation analysis is beyond the scope of this paper, we did want to get a sense of how work in CSCW builds and maintains ties to other CSCW researchers by drawing on CSCW publications. To that end, we decided to conduct an informal analysis of the citation patterns from recent CSCW conferences as derived from the records of papers presented at CSCW 2000 and CSCW 2002, contained in the Association for Computing Machinery's Digital Library [2]. Each record contains the list of entries from the paper's reference section, extracted using OCR software. We manually coded each reference for its source and year. This task had to be conducted by hand due to OCR errors and idiosyncrasies in the way authors formatted their reference sections – constraints that pose a challenge to researchers interested in conducting a full-fledged citation analysis of the field.

In this analysis we were interested in asking two main questions: First, from what sources do CSCW authors draw their references, and second, what is the impact of publications over time. We expected to use data on where and when cited documents were published to trace the lineage of ideas in the field. It should be noted that the citations serve a variety of purposes, and are not necessarily indexes of the importance or relevance of cited works (e.g., [20, 23]), although it has been noted that citations are better indicators of intellectual rather than social ties [29].

3.3.1 Citations by publication source

A breakdown of the publication sources of cited works is shown in Figure 4. As one would expect, authors look to journals for a large proportion of their references, and that the CSCW conferences (i.e., CSCW, ECSCW, and GROUP) also provide a large number of cited works. In addition, CHI is the second most cited conference behind CSCW itself. Perhaps the most interesting implication from these data relates to the fact that the universe of published CSCW papers is significantly smaller than any of the other categories[1], yet authors draw a significant proportion of their citations from that pool.

3.3.2 Citations by publication date

As we mentioned above, there is a good deal of turnover in the composition of the central core of CSCW researchers as new authors begin publishing in the field, supplanting some existing authors. One question that we may ask is whether older papers retain their influence in the field. As Figure 5 indicates, there is a strong bias toward citing recent papers within CSCW. In fact, 75% of cited works were published within the last eight years. One implication is that as individuals leave the field, references to their work do not persist. This is not to say that works that are no longer cited are not influential or that their authors no longer

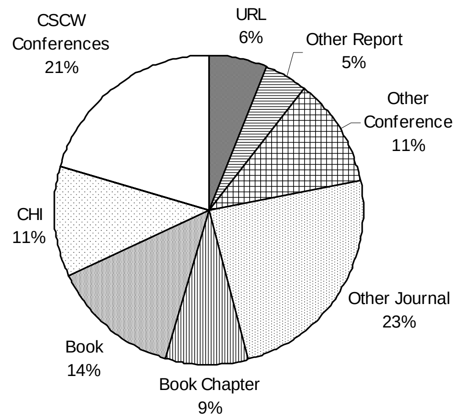

Figure 4. Breakdown of publication venues of works cited in the CSCW 2000 and CSCW 2002 conference proceedings. The category "CSCW Conferences" includes CSCW, ECSCW, and GROUP. The category "Other Report" includes dissertations, technical reports, unpublished

impact the field. There are a variety of reasons that a still influential paper may not be cited, and there are many ways to influence the field that do not leave traces in the citation network.

Over time the ideas presented in a particularly influential paper may be incorporated and extended by later papers (which consequently become the authoritative citations), or may become implicit in the thinking of members of the field. As an example, with the evolution of the worldwide web, authors may be less likely to cite Vannevar Bush's visionary article "As We May Think" [7][2]. In addition, with much of the work in our field inexorably intertwined with the progress of technology, many of the precursors to current research are fairly recent when compared with research in the sciences, and certainly with the humanities.

4. DISCUSSION

This paper used analysis of coauthorship networks in HCI and CSCW to answer a number of questions about the evolution and structure of the CSCW research community. In terms of the

[1] As an example, the proceedings from a typical CSCW conference include 40-50 papers, while the CHI 2004 proceedings included the text of nearly 200 papers and short-papers. In addition, CHI occurs every year, while CSCW only takes place in alternate years (of course only a subset of CHI papers are directly relevant to CSCW research). The number of potentially relevant journal articles and books is difficult to estimate, but is assuredly very large.

[2] It has not been lost on us that the coauthorship and citation records of this paper will affect future similar analyses of the field.

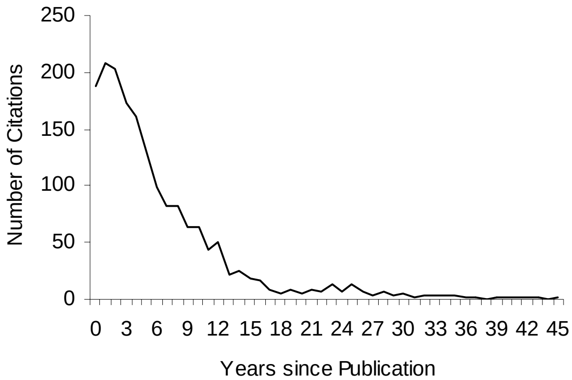

Figure 5. The distribution of the age of works cited in the CSCW 2000 and CSCW 2002 conference proceedings.

specific aims from the introduction, the results suggest the following conclusions.

First, with respect to the ties between the CSCW community and the larger HCI community, CSCW researchers have maintained a steady association with the HCI world. That is, during the period when CSCW emerged as a separate sub-field, CSCW researchers had roughly equal numbers of CSCW coauthors and HCI coauthors. This ratio has held constant through the present, suggesting that there continues to be a vital association between the CSCW community and the larger HCI community, at least as measured in terms of coauthorship. In terms of citation patterns, however, there is a more exclusive focus on the CSCW community. For example, in an examination of CSCW 2000 and 2002 publications, 21% of all citations were to CSCW works – compared to smaller percentages from much larger sources, such as 11% of all citations to the CHI proceedings.

Second, with respect to size and composition of the CSCW community over time, the community appears to be shrinking and has replaced itself almost completely over the preceding decade. Specifically, when viewed in terms of the giant component, the size of the CSCW coauthorship network peaked at 294 researchers during the five-year period 1996-2000, and has since declined to 188 researchers. In addition, the composition of the CSCW coauthorship network has experienced nearly complete turnover during the period 1988 through 2002. That is, of the 109 researchers in the coauthorship network in the interval of 1988-1992, only sixteen remained active in the interval of 1998-20002. Finally, even within the same cohort – i.e., those researchers who appeared for the first time in the CSCW giant component during the same period – there was high volatility. For instance, in the 1988-1992 cohort there was a 50% turnover in the composition of the top ten researchers by centrality over a ten-year period. Similarly, in the 1993-97 cohort there was a 70% turnover in the top ten over a five-year period. Analysis of cited works in CSCW publications suggests a possible association between these observed turnover rates – and the lack of visibility (in terms of citation) for older CSCW articles.

Third, with respect to the visibility of CSCW researchers within the HCI community, researchers central to the CSCW community tended to be central within the HCI community. For example, five of the twenty most central people in the CSCW cumulative network over the period 1982-2003 were also among the twenty most central people in the HCI cumulative network. Similarly, when ordered by percentile on centrality, highly central CSCW researchers tended to be central HCI researchers. However, there were a number of central HCI researchers with low centrality in the CSCW community. Also, there were a small number of cases where researchers had relatively high centrality in the CSCW community (e.g., in the top quartile) and relatively low centrality in the HCI community (e.g., in the bottom half).

4.1 Limitations of the findings

The findings reported in this paper represent an initial attempt to understand the ecology of collaboration within the field of CSCW. Consistent with efforts in other disciplines (e.g., physics) we focused on coauthorship networks. This approach has a number of advantages, notably access to a robust longitudinal set of data through the hcibib database. However, there are some weaknesses in choosing coauthorship as the metric of collaboration.

First, coauthorship only captures one kind of activity within a disciplinary community. That is, we noted high turnover in the composition of the coauthorship giant component, but this is not the same thing as high turnover in the CSCW field more broadly. For example, authoring papers may be a contribution uniquely associated with a certain part of a researcher's development. In other words, people may enter a field long before they coauthor their first paper through students volunteering. Similarly, people may stay in a field for some time after their peak period of coauthorship and perform other valuable functions, such as mentoring, reviewing, chairing conferences, and serving in administrative roles. Therefore, while providing a solid picture of one very important kind of contribution, coauthored scholarly papers, coauthorship networks can miss other critically important behaviors and contributions. In addition, our decision to use only the hcibib database leads to a network that misses potentially relevant publications from outside venues (e.g., a CSCW oriented paper in a Psychology journal).

A second issue with coauthorship is that centrality within a coauthorship network may be less a marker of visibility and influence within a field, and more a marker of a particular collaboration style. For instance, consider a researcher who brings together other researchers, yet does not share coauthorship in any resulting papers. This role of "matchmaker" is obviously critical, but in terms of coauthorship data, the matchmaker's activity is invisible (or may only be recorded in the acknowledgements – which don't create a trace in databases like hcibib). Contrast the matchmaker role with the behavior of the "bridging" role, where instead of making a coauthorship match – the bridging person coauthors separate papers with each potential collaborator – which creates a coauthorship trace, and therefore boosts centrality. We don't believe that individual researchers adopt deliberate bridging strategies, but we do wonder whether there are particular organizational structures and incentive schemes that reward bridging behavior. For example, within corporate research labs there may be more pressure to collaborate with peer professional researchers who represent different domains of knowledge, which would encourage bridging versus matchmaking.

4.2 Significance of the findings

The results reported here are significant for several reasons. First, while CSCW has been concerned with detailed analyses of collaboration, there has been less attention to ecological aspects of collaboration. That is, what is the larger structure of collaboration? What is the impact of particular structures? How do structures change over time? We examined these questions in the context of CSCW research and uncovered several features of the CSCW world that surprised us, even though we are CSCW researchers. In particular, we did not expect to find such high volatility in terms of the composition of the coauthorship network over time. From one perspective, the high turnover could be interpreted as a healthy indicator. That is, new people enter the field, presumably bringing new ideas, and therefore refresh the field. However, another interpretation could be that high turnover reflects churn due to lack of consensus about core questions.

A second significant aspect of the findings is the continuing strong association between CSCW research and the larger world of HCI research. That is, at least in terms of coauthorship, people in the CSCW community continue to work closely with people in the broader HCI community. This finding surprised us, at least to the extent that we perceived a greater independence of the CSCW community. For instance, the emphasis on ethnomethodology in the CSCW literature could be seen as an indictment of HCI, at least to the extent that traditional HCI research reflects a strong foundation in psychology, cognitive science, and computer science. Either CSCW is not as distinctive as we imagined, and remains strongly defined by the roots in traditional HCI – or CSCW has transformed HCI – and the continuing association of HCI and CSCW research reflects this transformation. The pattern of citation in CSCW publications, (i.e., more focused on the CSCW literature), is more consistent with a depiction of the CSCW community as distinct from the larger HCI community. Further research should address this apparent inconsistency.

Thirdly, the ongoing centrality of a small set of researchers over time – with periodic small additions to this core – suggests the vulnerability of ecologies of collaboration, such as the field of CSCW. That is, while the CSCW community appears fairly vital on key dimensions, such as ongoing conferences and growth in number of publications, there may be a larger than expected dependence on a small sub-set of people to maintain this vitality. Specifically, if a central group represents a research and a demographic cohort, there may be some question about the sustainability of the field if a significant fraction of the central cohort leaves the field in a short period of time, such as through retirement. At an intellectual level, the continued centrality of a small group of researchers may dampen diversity. In particular, it is possible to imagine a scenario where limited publication opportunities combined with a small group of central researchers results in a kind of "lockout" where there is no room for new work. The decline in the size of the CSCW giant component over the past few years along with reduced attendance at key conferences (e.g., CSCW) – may be reflections of this lockout condition. However, other factors certainly play a role, including the recent economic downturn and the diminished importance of corporate research labs.

4.3 Next steps

This preliminary research on the structure of coauthorship networks within CSCW suggests a number of next steps in terms of theoretical development and in terms of system design and deployment.

4.3.1 Theory

One critical question that the CSCW coauthorship network analyses cannot fully answer concerns the origin and maintenance of collaborations. Databases such as hcibib represent retrospective traces of collaboration successes, at least as measured by resulting publications. It would be interesting to know what made these collaborations successful. Specifically, we know that geographical proximity strongly affects collaboration (e.g., [19]). We propose to explore the function of collocation in collaboration outcomes by using data on past collaborations, such as contained in hcibib, to prime researchers to answer questions about the formation and maintenance of collaborations. In particular, we are interested in the physical proximity of collaborators during collaboration, or their physical distance, as well as their relative social status, or their social distance. Initial examination of the top collaborating dyads from the CSCW network suggests that, at least in terms of physical distance, collaboration seems to depend heavily on collocation – since only four of the top collaborating dyads appear to have initiated and sustained collaborations at a distance. To further explore the impact of physical and social distance, we have developed a Web-based questionnaire mechanism that uses data from a coauthorship database to prompt people about past collaborations (similar to a survey described in [19]). For instance, on a sample of published coauthored papers, a respondent will be asked to assess the intellectual impact of a paper, physical proximity to each author, and social status relative to each author. Through these data we plan to develop a clearer picture of the antecedents of collaboration, and the relationship of these factors to collaboration outcomes – and to individual outcomes, such as increased network centrality.

4.3.2 Systems

There are a number of documented efforts in the literature that describe either direct or indirect use of social networks primarily in service of finding information (e.g., Answer Garden [1]) or in service of finding who might have information (e.g., ReferralWeb [18], iKnow [9]). In addition, recent activity in the private sector (e.g., Friendster.com and Orkut.com) suggests a popular interest in using social networks, such as to make new friends. We believe that coauthorship networks represent a rich source of data for driving existing and future applications based on social networks. In particular, the disambiguated hcibib data used for this paper will be made available for testing and development of social networking aids. In addition, we believe coauthorship data will be valuable in identifying gaps – such as synergistic research programs that would benefit from collaboration – as well as what we have come to term "binding sites." That is, in examination of the CSCW network over time, we have observed the emergence of sub-groups (such as the stream of research on Lotus Notes in the early 1990s) that represent potential ties to larger communities of research, such as management information systems departments within business schools, that appear to come and go – without any elaboration. Awareness of these emerging ties, which have the potential to introduce new researchers and new topics, may increase with easy visualization and analysis. One result may be greater vitality within the CSCW field and overlapping relationships that are as rich and important as the current association of the CSCW community with the larger HCI community.

5. ACKNOWLEDGMENTS

We would like to thank Mark Newman, Mark Handel, and our anonymous reviewers for their advice and suggestions on this work, and Gary Perlman for maintaining the hcibib database and permitting us to analyze this data source. We would also like to acknowledge Jonathan Grudin's good humor regarding the title of this paper.

6. REFERENCES

[1] Ackerman, M. S. Augmenting Organizational Memory: A Field Study of Answer Garden. *ACM Transactions on Information Systems (TOIS)*, *16*, 3, (1998), 203-224.

[2] ***The Association for Computing Machinery Digital*** *Library.* http://www.acm.org/dl

[3] Bannon, L. J. Discovering CSCW in *Proceedings of 15th Information Systems Research in Scandinavia Seminar* (Larkollen, Norway, August 9-12 1992)

[4] Barabasi, A.-L., Jeong, H., Ravasz, E., Neda, Z., Schuberts, A. and Vicsek, T. Evolution of the Social Network of Scientific Collaborations. *Physica A 311*, (2002), 590-614.

[5] Batagelj, V. & Mrvar, A. Pajek – Analysis and visualization of large networks. In M. Junger & P. Mutzel (Eds.), *Graph drawing software* (2003), 77-103. Berlin: Springer.

[6] Borgman, C. L. & Siegfried, S. L. Getty's Synoname and Its Cousins: A Survey of Applications of Personal Name Matching Algorithms. *Journal of the American Society for Information Science 43*, 7 (1992), 459-476.

[7] Bush, V. As We May Think. *Atlantic Monthly 176*, 1 (1945), 101-108.

[8] Cheswick, B., Burch, H. and Branigan, S. Mapping and Visualizing the Internet. *Proceedings of the 2000 USENIX Annual Technical Conference* (San Diego, CA, June 18-23, 2000).

[9] Contractor N., Zink, D. and Chan, M. IKNOW: A tool to assist and study the creation, maintenance, and dissolution of knowledge networks. In Toru Ishida (Ed.), *Community Computing and Support Systems, Lecture Notes in Computer Science 1519* (1998), 201-217. Berlin: Springer-Verlag.

[10] Dodds, P. S., Muhamad, R., & Watts, D. J. An experimental study of search in global social networks. *Science*, *301*, 5634 (August 8, 2003), 827-829.

[11] Duderstadt, J. J. *Preparing for the Revolution: Information Technology and the Future of the Research University*. National Academies Press, Washington, DC, 2002.

[12] Finholt, T. A. and Teasley, S. D. The Need for Psychology in Research on Computer Supported Cooperative Work. *Social Science Computer Review 16, 1* (1998), 40-52.

[13] Grossman, J. *The Erdös Number Project*. http://personalwebs.oakland.edu/~grossman/erdoshp.html

[14] Grudin, J. CSCW: The Convergence of Two Development Contexts in *Proceedings of ACM Conference on Human Factors in Computing Systems* (New Orleans, LA, 1991) ACM Press, 91- 97.

[15] Grudin, J. Computer-Supported Cooperative Work: History and Focus. *IEEE Computer 27, 5* (1994), 19-26.

[16] Hickey, T. B. *Development of a Probabilistic Author Search and Matching Technique for Retrieval and Creation of Bibliographic Records.* Report No. OCLC/OPR/RR-81/2. OCLC Office of Planning and Research, Dublin, OH, 1981.

[17] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. and Barabasi, A-L. The Large-Scale Organization of Metabolic Networks. *Nature, 407* (2000), 651.

[18] Kautz, H., Selman, B. and Shah, M. The Hidden Web. *The AI Magazine*, *18*, 2 (1997), 27-36.

[19] Kraut, R. E., Egido, C. and Galegher, J. Patterns & Contact and Communication in Scientific Research Collaboration. In Jolene Calegher & Robert E. Kraut (Eds.). *Intellectual teamwork: Social and technological foundations of group work* (1990). NJ: Lawrence Erlbaum Association.

[20] Leydesdorff, L. & Amsterdamska, O. Dimensions of citation analysis. *Science, Technology and Human Values*, *15*, (1990), 305-335.

[21] Milgram, S. The Small World Problem. *Psychology Today 2*, (1967), 60-67.

[22] Moody, J. Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology 107*, (2001), 679-716.

[23] Moravcsik, M. J. & Murugesan, P. Some Results on the Function and Quality of Citations. *Social Studies of Science 5*, (1975), 86-92.

[24] Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Review 45*, (2003), 167-256.

[25] Newman, M. E. J. Who Is the Best Connected Scientist? A Study of Scientific Coauthorship Networks. *Physics Review 64, 4* (2001), 046132-1.

[26] *The Oracle of Bacon*. http://www.cs.virginia.edu/oracle/

[27] Perlman, G. The HCI Bibliography: Ten Years Old, but What's it Done for Me Lately? *Interactions 6*, 2 (1999), 32-35.

[28] Wasserman, S. and Galaskiewicz, J. *Advances in Social Network Analysis*. Sage Publications, Thousand Oaks, CA, 1994.

[29] White, H. D., Wellman, B., & Nazer, N. Does Citation Reflect Social Structure? Longitudinal Evidence From the "Globenet" Interdisciplinary Research Group. *Journal of the American Society for Information Science and Technology 55*, 2 (2004), 111-126.

[30] http://www.crew.umich.edu/hcinet/