LA-UR- 09-06291

Approved for public release; distribution is unlimited.

Title: An in-depth longitudinal analysis of mixing patterns in a small scientific collaboration network

Author(s): Marko A. Rodriguez, T-5/CNLS
Alberto Pepe, UCLA

Intended for: Submission to Scientometrics (Springer) Journal
ISSN: 0138-9130

Los Alamos
NATIONAL LABORATORY
EST. 1943

An in-depth longitudinal analysis of mixing patterns in a small scientific collaboration network

Alberto Pepe* Marko A. Rodriguez†

29 August 2009

Abstract

Many investigations of scientific collaboration are based on large-scale statistical analyses of networks constructed from bibliographic repositories. These investigations often rely on a wealth of bibliographic data, but very little or no other information about the individuals in the network, and thus, fail to illustate the broader social and academic landscape in which collaboration takes place. In this article, we perform an in-depth longitudinal analysis of a small-scale network of scientific collaboration ($N = 291$) constructed from the bibliographic record of a research center involved in the development and application of sensor network technologies. We perform a preliminary analysis of selected structural properties of the network, computing its range, configuration and topology. We then support our preliminary statistical analysis with an in-depth temporal investigation of the assortativity mixing of these node characteristics: academic department, affiliation, position, and country of origin of the individuals in the network. Our qualitative analysis of mixing patterns offers clues as to the nature of the scientific community being modeled in relation to its organizational, disciplinary, institutional, and international arrangements of collaboration.

Keywords. Scientific collaboration networks – sensor network and wireless research – small-scale networks – network topology – network evolution – mixing patterns – discrete assortativity – homophily.

*Alberto Pepe. Center for Embedded Networked Sensing. University of California, Los Angeles. Email: apepe@ucla.edu

†Marko A. Rodriguez. T-5, Center for Nonlinear Studies. Los Alamos National Laboratory. Email: marko@lanl.gov

1 Introduction

Scientific communities have large, well-established, and relatively well structured digital footprints which have increasingly been the focus of specialized research. These footprints, composed of scholarly publications and related artifacts, have been employed for bibliometric analyses involving coauthorship, citation, co-citation, acknowledgments, and other such indicators of scientific productivity and knowledge production [7]. Furthermore, as the majority of scientific publications are now available and consulted online, a number of recent analyses have progressively studied not only the production of scholarly artifacts but also their usage. Recent studies include: an analysis of the structure of readership networks in the field of education research [9] and a large-scale analysis of user request logs (clickstreams) gathered from scholarly web portals [4]. In general, these kinds of research are made possible because the scientific community is a large social institution that is open to investigation.

Coauthorship patterns are perhaps the most studied scholarly and scientific phenomena. Recent studies of coauthorship have analyzed the literature production within specific domains such as high energy physics [23], genetic programming [35], neuroscience [8], nanoscience [32], library science [20], economics [17], organizational science [1], and psychology and philosophy [13]. Similar analyses have also been comparative in nature and have explored social and normative differences of coauthorship behavior across different domains [30]. Moreover, an increasing number of studies of this kind have accounted for the evolving component of scientific collaboration [2, 10, 37, 33].

More specifically, coauthorship patterns have been widely and actively studied from a social network analysis perspective for over two decades [19, 18, 14]. Most social network research involved with coauthorship is based upon this underlying concept: two individuals (nodes) are regarded as coauthors if they appear together in the author list of a publication (edge). This relational structure works reasonably well when investigating coauthorship patterns in scholarly arrangements where publications are authored by relatively small groups. It is true that some scientific domains have experienced an increase in the number of authors per publication making it impossible to discern the nature and extent of individual contributions to a publication [12]. A striking example of this phenomenon can be found in the domain of high-energy physics where author lists for a single publication often include tens or even hundreds of authoring researchers [36]. For this reason, a number of recent studies of physics collaboration supplement traditional analytic techniques with more qualitative methods of survey research, i.e., directly asking authors to indicate the real nature of their contributions to a publication [34, 3]. However, besides the singular case of high-energy physics, the vast majority of scholarly coauthorship networks incorporate collaboration circles of only a handful of authors per publication, suggesting that coauthorship activity can be adequately employed to construct a valid social network of collaboration [25].

In this article, we perform a temporal analysis of a coauthorship network constructed from the bibliographic record of a research center involved in the development and application of sensor network and wireless technologies. The network studied here is relatively small in size (in its largest year, $N = 291$), compared to networks generally analyzed in related research. The small size of the collaboration network results in a fundamental advantage: besides analyzing certain large-scale structural properties of the network, we can explore the social and academic arrangements in which collaboration patterns evolve, based on a set of node characteristics. The study of node characteristics can provide insights into the level of homophily in a network, i.e. the extent by which individuals with given characteristics create ties with others with similar characteristics [22]. In analytic terms, homophily can be studied by computing a network's assortative mixing, i.e. the tendency for network's nodes to preferentially connect with similar nodes [28]. In this article, we perform a longitudinal analysis of assortative mixing based on academic department, affiliation, position, and country of origin of the individuals in the collaboration network under study.

Studying networks in terms of specific characteristics of their constituent nodes is not a new technique. Research interests [16], academic domain [24], geographical location [21], age group [5], and country of origin [31] are examples of node characteristics that have been investigated in previous analyses of scholarly collaboration. The present study, exploratory in nature, differs from previous ones for it tries to tie the quantitative analysis of the network's structural properties to a qualitative explanation of the social and academic landscape in which the network is based. In other words, after performing a quantitative analysis of the scientific collaboration network, we were interested in finding out whether the network's structural properties were driven by factors exogenous to its representation. We found that certain social and academic dynamics, for example the emergence of new international collaborations or the inception of new interdepartmental efforts, had varying levels of control in the resulting topology and configuration of the scientific collaboration network.

The present article is organized as follows. In the next section, we present our data and perform a preliminary analysis of the network's range, configuration and topology over time. In the following section, we support our preliminary analysis with an in-depth qualitative analysis of the observed mixing patterns of a set of node characteristics. We conclude this article by noting that our in-depth longitudinal analysis of mixing patterns can offer clues as to the nature of the scientific community being modeled in relation to its organizational, disciplinary, institutional, and international arrangements. Furthermore, it can provide a sociological explanation of the observed collaboration patterns without the need to rely on surveys and similar ethnographic techniques.

2 Present study

This article presents the findings of a study of scientific collaboration at the Center for Embedded Networked Sensing (CENS).[1] CENS is a National Science Foundation (NSF) Science & Technology Center funded in 2002. CENS supports interdisciplinary collaborations among faculty, students, and staff of five partner universities in Southern California: University of California, Los Angeles (UCLA); University of Southern California (USC); University of California, Riverside; California Institute of Technology (Caltech); and University of California, Merced. From 2005, CENS features a headquarter office base located within Boelter Hall at UCLA.

The mission of CENS research is to use sensor network systems to reveal previously unobservable phenomena. From its inception, CENS has developed and deployed sensor network devices for the study of a wide range of natural phenomena, such as seismic activity, fluid contaminant transport, and bird breeding behavior. Besides these pursuits in the natural sciences, the social and built environments have progressively become the focus of applied CENS research: sensing mobile systems are being employed for the study of public health, environmental protection, urban planning, and cultural expression.

The type of research conducted at CENS now spans a wide spectrum of disciplines and applications (from biology to seismology, from wireless telecommunications to statistics, from education to environmental science) requiring continuous cooperation among individuals that, otherwise, would probably not interact beyond the walls of traditional university departments and faculties. In such a diverse scholarly and scientific landscape, distributed collaboration on multi-disciplinary subjects constitute a fundamental leverage for scientific research.

2.1 Data collection

Computing the range, population and configuration of an interdisciplinary, multi-institutional research center like CENS can be an arduous task.

> "In order to be recognisable as such, a system must be bounded in some way. However, as soon as one tries to be specific about the boundaries of a system, a number of difficulties become apparent" [11, p. 139].

In the case presented in this article, these difficulties have to do with the inherently open and dynamic nature of modern science research centers. Unlike other types of organizational arrangements for which the boundaries are more or

[1] The website of the Center for Embedded Networked Sensing (CENS) is available online at http://research.cens.ucla.edu/

less evident (e.g. academic institutions and departments, corporate and government centers), many modern research centers and laboratories act as umbrella organizations with very flexible and blurry boundaries. CENS, for example, includes researchers from multiple institutions and disciplines. CENS scholars seamlessly interact with each other within and beyond institutional and departmental boundaries: collaboration patterns are ubiquitous and non-uniformly distributed. Researchers affiliated with CENS may also be affiliated with other research laboratories and perform interdisciplinary work on other projects and under different affiliations. Moreover, many CENS collaborations include researchers that are not officially affiliated with CENS. In other words, the nature and extent of contribution to CENS collaboration depends on a number of economical, political, and scholarly factors, and is not solely restricted to individuals officially affiliated with CENS. Under these conditions, what is the best way to construct a network that accurately captures scientific collaboration of this research center?

Previous environment-specific studies of collaboration delineate the population under study by relying on publication data contained in an institutional repository [1] or domain-specific bibliographic databases [20] to mine patterns of coauthorship that take place within a given institution or academic domain, respectively. For the purpose of this article, we used a similar mechanism, thus including in our population not only CENS members, but also individuals that though not officially affiliated, have contributed to the production of CENS or CENS-related scholarly publications.

We constructed a database of publications by assembling the items included yearly in the NSF Annual Reports, which contain the official list of documents published by CENS for a given fiscal year.[2] For the available reporting period (2002 through 2008) and for every publication, we collected the following information: a) author names, b) publication title, c) publication type, d) publication venue, and e) publication date. For the purpose of this article, we utilized author names and publication dates (years) to construct a coauthorship network, i.e. a network consisting of individuals (nodes) that are connected to each other (via edges) if they are recorded as authors on the same scholarly publication. Data were initially collected in the BibTeX format.[3] The resulting bibliographic dataset consisted of 547 manuscripts (370 conference proceedings, 159 journal articles, 17 book chapters and 1 book), published over a period of 7 years (2001–2007).

This bibliographic database was used to generate a weighted undirected network in which nodes represent authors and edges represent coauthorship activity among them. For example, if the present paper had to be included in this network, its authors (Pepe and Rodriguez) would become two distinct nodes, connected by an edge. In order to determine the weights between nodes, i.e.

[2] CENS Annual Reports are available online at http://research.cens.ucla.edu/about/annual_reports/

[3] BibTeX available online at http://www.bibtex.org/

the strength of collaboration among coauthors, we used a weighting mechanism proposed by Newman [26] by which the weight of the edge between nodes i and j is:

$$w_{ij} = \sum_k \frac{\delta_i^k \delta_j^k}{n_k - 1}, \tag{1}$$

where δ_i^k is 1 if author i collaborated on paper k (and zero otherwise) and n_k is the number of coauthors of paper k. For the example above, the edge between authors Pepe and Rodriguez would have $w_{ij} = 1$. An article written by three authors (e.g., Pepe, Rodriguez, and Bollen) would result in three edges (Pepe-Rodriguez, Pepe-Bollen, and Rodriguez-Bollen), each one with $w_{ij} = 0.5$. And so on. As such, this weighting mechanism confers more weight to small and frequent collaborations, based on the assumptions that: i) publications authored by a small number of individuals involve stronger interpersonal collaboration than multi-authored publications, and ii) authors that have authored multiple papers together know each other better on average and thus collaborate more strongly than occasional coauthors [26].

The resulting network data were "sliced" according to publication year yielding to 7 separate networks, each one representing the cumulative collaborative effort of CENS researchers up to that year. These networks of coauthorship for years from 2002 to 2007 are depicted in Figure 1 (a through f).[4] In all depicted networks, the diameter of each node is defined by its eigenvector centrality [6] and the darker gray nodes denote individuals that are primarily affiliated with CENS.

2.2 Preliminary analysis of network evolution: range, configuration, and topology

For each one of the coauthorship networks under study, for years 2001 through 2007, we calculated some fundamental network statistics, presented in Table 1.

An analysis of these statistics provides insights into the evolution of the CENS coauthorship network over time. The first three rows of the table contain, for every year in the period under study, the cumulative number of a) authors, i.e. nodes in the network, b) publications (journal articles, conference papers, etc.), and c) collaborations (coauthoring events), i.e. edges in the network. When analyzed over time, these three values all follow a similar trend, which highlights two distinct time periods: a first term (2001-2004) during which all values increase sharply (roughly doubling in size from year to year), and a second term (2004-2007), during which the growth slows down. In particular the author count values indicate that CENS quickly became large and diversified in its population in the first term reaching a solid population base of collaborators by the year 2004. In the second term, from 2004 to 2007, the author population

[4]The network depicting year 2001, though analyzed in this article, is not diagrammed as it is very sparse. Its visualization does not inform the present dicussion.

(a) 2002 (b) 2003

(c) 2004 (d) 2005

(e) 2006 (f) 2007

Figure 1: The CENS coauthorship network under study sliced according to year of publication (2002 through 2008, cumulative data). Vertex diameter represents the eigenvector centrality score of the vertex, where more central vertices have larger diameters. Moreover, the darker gray nodes denote individuals that are primarily affiliated with CENS.

quantity/year	2001	2002	2003	2004	2005	2006	2007
Authors (nodes)	35	68	127	203	228	278	291
Publications	23	69	175	303	418	496	547
Collaborations (edges)	182	346	690	1248	1598	2158	2536
Connected components	4	5	8	13	9	6	5
Diameter	3	4	6	7	8	8	7
Average path length, ℓ	1.543	2.324	3.090	3.038	3.339	3.224	2.944
Clustering coefficient, C	0.645	0.543	0.432	0.387	0.389	0.337	0.329
Degree assortativity, r	0.432	0.299	0.272	0.187	0.180	0.166	0.165

Table 1: A summary of the fundamental network statistics of the CENS coauthorship network for the time period 2001-2007: number of authors, publications and collaborations (range), number of connected components and diameter of the largest connected component (configuration), and average path length, clustering coefficient and degree assortativity (topology). All values presented are cumulative.

increased again, but to a much lesser extent (from 203 to 291 individuals), while the number of published works and collaborations maintained a regular growth (from 1248 to 2536 collaborations), suggesting the formation of a core CENS authoring base.

This finding is confirmed by a quick analysis of the network's configuration. The number of connected components, i.e. the number of maximal connected subgraphs, goes from 4, in 2001, to 13, in 2004, indicating that the network becomes more fragmented in the first term, even if collaboration is overall increasing. In the second term, however, the number of connected components drops and the network quickly solidifies into a giant component, which indicates a solid base of strong collaboration. By looking at the the network diameter, i.e. the length of the longest geodesic path in the largest connected component, the formation of the giant component in year 2004 becomes evident. This is further reinforced by a quick analysis of Table 2, which lists component populations by year.

year	#	population
2001	4	18 6 7 4
2002	5	38 16 8 4 2
2003	8	93 10 7 5 4 3 3 2
2004	13	150 10 8 8 5 4 3 3 3 3 2 2 2
2005	9	205 5 3 3 3 3 2 2 2
2006	6	262 5 4 3 2 2
2007	5	280 4 3 2 2

Table 2: Node populations of connected components in the CENS coauthorship network, by year.

The preliminary analysis of these first two sets of values from Table 1 gives us a good understanding of the evolution of the range and configuration of the network over time. A third set of values, presented in Table 1 (average path length, clustering coefficient, and degree assortativity) can be investigated to provide an in depth understanding of its topology.

The average path length is the average length of the shortest paths between all possible node pairs and, in turn, an indicator of the efficiency of information transfer in a social network [38]. Short average path length, and thus high information transmission, are typical characteristics of many real and small-world networks [39]. In the network under study, the average path length is about 1.5 in 2001; it grows steadily in the first term, reaching a value of about 3.0, which stays roughly constant throughout the second term. This indicates that once the CENS authoring base is formed, an average of 3 steps are necessary to transfer information among any two pairs of nodes. Remarkably, this value resembles more closely that found in typical small-world networks, such as movie actors ($\ell = 3.65$) [39], than that of scholarly coauthorship networks, such as mathematics ($\ell = 9.5$) and neuroscience ($\ell = 6$) [2]; yet, this relatively low average path is possibly due to the relatively small size of the network analyzed (the mathematics and neuroscience coauthorship networks have $N = 70,975$ and $209,293$, respectively).

The clustering coefficient measures the density of clique-like triangles in a network. High clustering coefficient coupled with short average path length indicates that a network exhibits small-world properties [29]. In the CENS coauthorship network, the clustering coefficient decreases steadily over time, from an initial value of 0.645 in 2001, to 0.329 in 2007. This suggests that the network becomes less cliquish and collaboration patterns becomes more uniform across the network over time. This trend reveals that the CENS network initially matches the typical topology of highly-clustered disciplines such as physics ($C = 0.56$) and biology ($C = 0.6$) [29] but later drops to the values normally recorded in less cliquish domains, such as mathematics ($C = 0.34$) [15].

A final indicator of network topology presented here is degree assortativity. In general, assortativity can be defined as the tendency for individuals (nodes) in a social network to establish connections preferentially to other individuals with similar characteristics [27]. The most common measure of assortativity is computed based on the individuals' degree centrality. In the network presented here, degree-based assortativity decreases steadily over time from a value of 0.432 in 2001 to 0.165 in 2007. Interestingly, the decline of the degree assortativity measure follows very closely that of the clustering coefficient — the correlation between the two is $\rho = 0.964$ (p-value < 0.005). This means that as collaboration patterns in the network become more sparse and uniform (decreasing C), they also become more mixed (decreasing r), i.e. highly-connected individuals begin to collaborate with lowly-connected ones. In the next section, we extend the study of assortativity to a set of discrete characteristics, namely authors' academic department, affiliation, position, and country of origin. Analysis of these mixing patterns allows us to understand the homophilious and heterophilious components that contribute to the observed growth of the network.

3 Studying network evolution in terms of discrete node characteristics

The preliminary analysis of the CENS coauthorship network presented in the previous section reveals the following scenario. In 2001, the network of collaboration is small and very fragmented. During the first few years of activity, however, the CENS group grows significantly in the number of authors and collaborations. By the end of 2004, a solid base of collaborating authors (i.e. a giant component) is formed. In the analyzed network, small-world effects become less prominent over time; in particular, average distance between individuals becomes larger (increasing ℓ), and collaboration patterns in the network become more sparse (decreasing C) and more mixed (decreasing r).

Although our preliminary analysis presents a fairly comprehensive account of the range, configuration and topology of the studied network of scientific collaboration over time, we believe that it fails to provide a sociological explanation of the dynamics underlying the observed patterns. In particular, we were curious to explore further the correlation between clustering coefficient and assortativity. Our preliminary analysis indicates that there exists a solid link between these two patterns: a) the network becoming more sparse and uniform, and b) collaboration patterns becoming more mixed. However, this analysis is restricted to degree assortativity and thus ignores other mixing patterns that might have contributed to the decrease in network clustering over time. For this reason, we became interested in deepening our understanding of the sociological and academic context of the CENS collaboration network to identify specific patterns that might account for the observed clustering pattern. For example, can we speculate that the network becoming more sparse is indicative of higher interdisciplinary collaboration and/or higher collaboration across different institutions? In this context, the question that we would like to address is: what specific mixing patterns are accountable for the decrease in the network's clustering coefficient? In the remainder of this article, we extend the temporal analysis, presented in the previous section, to a set of node characteristics. Our aim is to justify the observed clustering pattern in terms of specific social and academic characteristics that are more telling of the sociological aspects of the group than what degree assortativity alone can elucidate.

3.1 Further data collection

We collected these additional metadata relative to each author in the network under study: a) academic affiliation, b) academic department, c) academic position, and d) country of origin.

We collected these metadata via manual techniques, i.e. gathering required information on the authors' personal web pages and consulting online directories from university and department web sites. It is worth noting that all the parameters collected (except for country of origin) are subject to change over time,

Academic affiliation	
University of California, Los Angeles (UCLA)	148
University of Southern California (USC)	66
Massachusetts Institute of Technology (MIT)	10
California Institute of Technology (Caltech)	8
University of California, Riverside (UC Riverside)	7
University of California, Berkeley (UC Berkeley)	7
University of California, Merced (UC Merced)	4
State University of New York at Stony Brook (SUNYSB)	3

Academic department	
Computer Science	113
Electrical Engineering	80
Civil Engineering	23
Biology	19
Information Sciencea	9
Environmental Science	7
Education	5
Marine Biology	4
Statistics	3
Linguistics	3

Academic position	
Graduate Student (Grad Student)	97
Staff / Research Associate (Staff)	67
Full Professor (Professor)	44
Postdoctoral Student (PostDoc)	21
Associate Professor (Assoc Prof)	21
Assistant Professor (Assistant Prof)	20
Undergraduate Student (Undergrad)	5
Lecturer (Lect)	3

Country of origin	
United States (USA)	120
India	33
China	24
Italy	10
South Korea (Korea)	9
Australia	5
Greece	4
Iran	3
Taiwan	3
Mexico	3

Table 3: Population counts (10 most recurring values) for the collected node properties: authors' academic affiliation, department, position and country of origin

even in the short timespan studied in this article. Researchers and scientists might change institution, department and position in a seven-year time period. For this reason, we consulted not only authors' personal web sites, but also their curriculum vitae and biographies to record changes in their academic affiliation, department and position. Curriculum vitae were also useful to collect authors' country of origin, which, for the purpose of this study, we consider to be the country in which individuals pursued their high-school education. Table 3 presents the frequency counts for the collected author metadata: authors' academic affiliation, department, position and country of origin.

A quick analysis of Table 3 reveals that the collaboration network studied here consists mostly of scholars from UCLA and USC from the departments of Computer Science and Electrical Engineering. The network consists of a large number of graduate students,[5] but researchers and professors at all levels make up a large portion of the collaboration network. Finally, the network is very international in its population, with about half of the authors being from the United States, about a quarter from India and China and the rest from a number of other countries worldwide.

3.2 Analysis of network evolution by discrete assortative mixing

The temporal analysis of degree assortativity, presented in the previous section, indicates the extent to which individuals in the network co-author preferentially to other individuals with similar degree centrality. Using the newly collected author metadata — academic affiliation, department, position and country of origin — we can extend our investigation of assortativity to compute mixing patterns based on these discrete parameters. In our case, all analyzed parameters are nominal and we can thus measure discrete assortativity coefficient, r, using the following formula [28]:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \qquad (2)$$

where e_{ij} is the fraction of edges in a network that connect a node of type i to one of type j, a_i is the fraction of edges that have a node of type i on the head of the edge, and b_i is the fraction of edges that have a node of type i on the tail of the edge. Finally, $r = -1$ when there is perfect disassortative mixing, $r = 0$ when there is no assortative mixing, and $r = 1$ when there is perfect assortative mixing. In other words, the discrete assortativity coefficient, r, indicates the level of homophily of the network for a certain parameter. For example, if r for academic affiliation is 1.0, this means that individuals in the network only write papers with other individuals with same institutional affiliation. In this kind

[5]Graduate students are students both at Ph.D. and Master level. Although these have been incorporated into a single category, Ph.D. students are the vast majority in this group (over 95%).

of network, there are no multi-institutional collaborations. On the other side of the spectrum, we can imagine a completely disassortative network ($r = -1$) in which every single collaboraton (i.e. paper) in the network is authored by individuals that belong to different institutions.

Table 4 presents the discrete assortativity coefficients for the network under study based on authors' academic affiliation, department, position and country of origin, calculated at seven temporal snapshots of the network (2001 through 2007). A visual representation of these values is also presented in the plot of Figure 2.

property/year	2001	2002	2003	2004	2005	2006	2007
Academic affiliation	0.438	0.448	0.501	0.533	0.550	0.584	0.544
Academic department	0.463	0.535	0.574	0.560	0.555	0.516	0.474
Academic position	0.177	0.192	0.188	0.184	0.182	0.177	0.177
Country of origin	0.245	0.286	0.369	0.350	0.347	0.352	0.367

Table 4: Discrete assortativity coefficients for years 2001 through 2007 based on authors' academic affiliation, department, position and country of origin

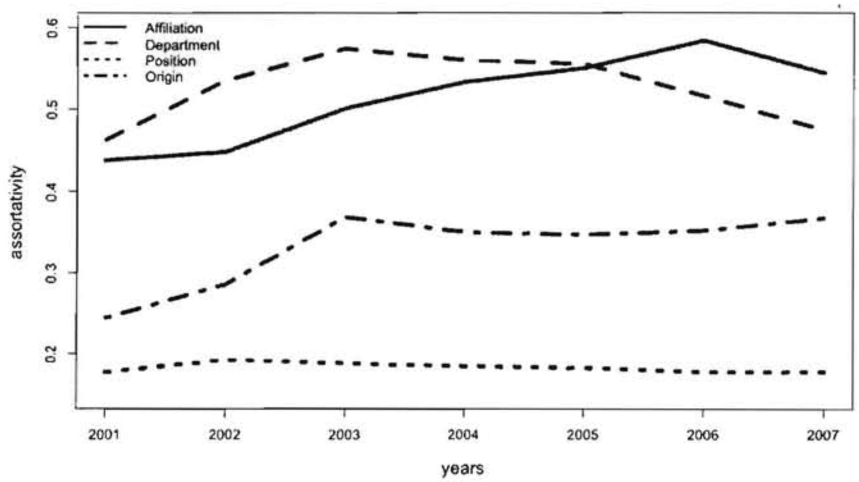

Figure 2: Plot of discrete assortativity coefficients for years 2001 through 2007 based on authors' academic affiliation, department, position and country of origin

Looking at Figure 2, it is easy to deduce that these assortativity coefficients are all within the same range — betweeen a minimum of 0.1 and a maximum of 0.6. Also, they do not change very much over the period under study — fluctuations in the 7-year period rarely exceed 0.1. Overall, the network is more assortative by academic affiliation and department, and less assortative by academic position and country of origin. A detailed temporal analysis of these coefficients can provide insights into the extent and nature of mixing patterns in the coauthorship network. For example, analyzing the assortativity by academic department, we can deduce how inter-disciplinary the CENS network is and how inter-discipliarity has changed over time. However, this analysis fails to reveal the specific components that contributed to these mixing patterns, e.g. *what collaborations are most responsible for the increase in inter-disciplinarity?*. In the remainder of this section we present, for each one of the studied characteristics, a detailed interpretation of our findings. Our aim is to push our understanding of the assortativity coefficients further, decomposing the observed collaboration patterns along specific components, to allow a more in-depth temporal analysis of the observed mixing patterns.

3.2.1 Academic affiliation

From Figure 2, the assortativity coefficient based on nodes' academic affiliation grows steadily over time, by about 0.1, from 0.438 in 2001 to 0.544 in 2007. This indicates that, overall, authors in the CENS coauthorship network have increasingly collaborated with other authors from the same institutional affiliation, in the time period under study. In the latest snapshot of the CENS network (year 2007) academic affiliation is the single most assortative characteristic, suggesting that CENS authors collaborate preferentially with individuals in their institution. This finding matches an earlier observation that the community structure of CENS collaboration matches very closely its institutional arrangement [31].

We would like to investigate this finding further, analyzing the specific intra- and inter-institutional collaborations that contributed to making the network more assortative over time. In order to do this, we inspect the most recurrent mixing patterns by affiliation in the last snapshot of the network, i.e., the institutional pairs that make up the majority of collaboration volume in 2007. These values are presented in Table 5. Clearly, as the coauthorship network under study is undirected, the order of the pairs (affil-1, affil-2) is not relevant.

The top four rows in Table 5 present the institutional pairs contributing to intra-institutional collaboration, whereas the bottom six rows present the pairs contributing to inter-institutional collaboration. So for example, in year 2001 there were 42 coauthorship activities among individuals affiliated with UCLA. From a network perspective, this means that in the year 2001, 42 edges, out of a total of 182 (from Table 1), were among nodes with parameter "UCLA". Collaboration between UCLA researchers increases steadily and reaches 1098 (out of a total of 2536) edges in 2007. In Figure 3, we plot the values of Table 5 as a fraction of the total volume of collaborations each year. Please note that the order of the institution pairs in Figure 3 is reversed with respect to Table 5, so that most prominent collaborations occupy the lower portion of the plot.

affil-1	affil-2	2001	2002	2003	2004	2005	2006	2007
UCLA	UCLA	42	112	262	468	678	948	1098
USC	USC	38	38	106	244	282	436	442
UCR	UCR	-	2	4	4	16	24	24
Caltech	Caltech	10	16	20	20	20	20	20
UCLA	USC	56	60	80	144	166	212	324
UCLA	MIT	8	10	20	44	50	64	78
UCLA	UCM	-	-	6	6	8	26	58
UCLA	Caltech	4	4	14	16	16	18	32
UCLA	UCR	6	6	6	8	16	20	30
UCLA	UCB	6	18	26	28	28	28	28

Table 5: Most recurrent academic affiliation pairs (top 10 results, cumulative values). Top 4 rows show intra-institutional collaboration. Bottom 6 rows show inter-institutional collaboration.

The stacked plot of Figure 3 allows us to decompose the assortativity coefficient trend lines of Figure 2 for discrete parameter: academic affiliation. From Figure 2, assortativty by affiliation increases steadily from 2001 to 2006 and finally drops slightly from 2006 to 2007. This trend can be understood in terms of the growth of intra- and inter-institutional collaborations, presented in Figure 3. From the stacked plot of Figure 3, we note that in 2001, the vast majority of recorded collaborations involve intra- and inter-institutional efforts between UCLA and USC individuals. In year 2007, the picture is not very different, with UCLA and USC still composing the bulk of the total volume of collaborations. However, a closer look at the components of the plot reveals that intra-institutional collaboration at UCLA has doubled in volume (from 0.2 to 0.4) while inter-institutional collaboration (UCLA-USC) has halved (0.3 to 0.15), compared to 2001 values. USC-USC collaboration stays roughly constant throughout the period under study. The increase in UCLA-UCLA and the decrease of UCLA-USC collaborations are the components that are most responsible for the increase in assortativity coefficient by affiliation from 2001 to 2007, presented in Figure 2.

There are some other collaboration dynamics that contribute to this trend. For example, besides UCLA-USC, the overall inter-institutional effort of UCLA decreases (e.g. collaborations with UC Berkeley and UC Riverside). Moreover, intra-institutional collaborations by Caltech researchers (which make almost 10% of the total volume in 2001) fade away over time. In sum, by year 2007, the collaboration scenario at CENS is largely dominated by publications authored within UCLA. Based on this finding, we can conclude that despite CENS's mission to be a multi-institutional research center, the temporal decomposition of coauthorship patterns demonstrates that CENS collaboration became less inter-institutional from 2001 to 2007 and consolidated around its main institution, UCLA. The steady increase in UCLA-UCLA collaboration can possibly be attributed to the construction of a CENS headquarter office at UCLA, completed in 2005. We can speculate that the CENS headquarter has brought UCLA scholars closer to each other, enabling interpersonal collaboration among them

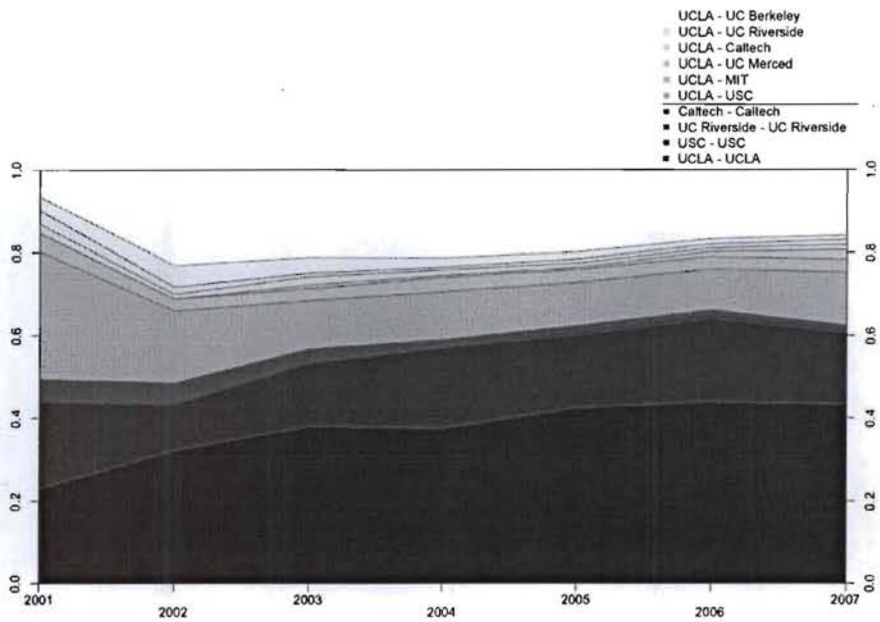

Figure 3: Top ten most recurrent academic affiliation pairs as fraction of total volume of collaboration. Darker polygons at the bottom are instra-institutional collaboration, while lighter polygons depict inter-institutional collaboration.

and, in turn, boosting coauthorship activity.[6]

3.2.2 Academic department

From Figure 2, the assortativity coefficient for academic department has the following trend. In year 2001, the CENS network is heavily assortative based on academic department ($r = 0.463$). In the following two years, assortativity increases even more, reaching a peak of 0.574 in 2003. This means that in 2003, the CENS coauthorship was very highly fragmented by department. By extension, we can speculate that at this time, collaboration patterns were vastly mono-disciplinary. In later years, however, assortativity by department decreases. Even though the value recorded in 2007 ($r = 0.474$) is roughly equivalent to the network's outset, the trend observed from 2003 to 2007 indicates the CENS collaboration network becoming more interdisciplinary. A decomposition of the observed coauthorship patterns, presented below, enables us to

[6]This is a fair assumption for the authoring of scientific conference papers, that have a much quicker publication turnaround than journal articles.

understand, more in depth, the extent of intra- and inter-departmental collaboration. For reasons of space, we do not include here the cumulative, yearly volume of collaboration among departments in the CENS network, as we did for the academic affiliation component, in the previous section. We limit ourselves to presenting, in Figure 4, a stacked plot that displays the most recurrent department pairs as fraction of total volume of collaboration.

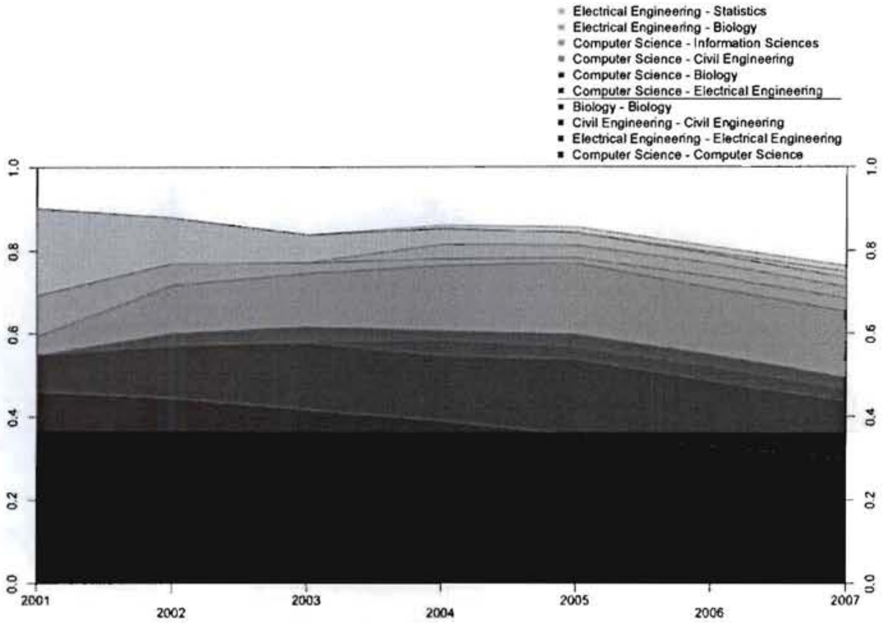

Figure 4: Top ten most recurrent academic department pairs as fraction of total volume of collaboration. Darker polygons at the bottom are intra-departmental collaboration, while lighter polygons depict inter-departmental collaboration.

In Figure 4, the darker, bottom four polygons are the department pairs that contribute the most to intra-departmental collaboration, whereas the top six polygons present the department pairs that account for most of inter-departmental collaboration. At the network's outset, CENS collaboration is dominated by intra-departmental collaborations in Computer Science and Electrical Engineering. The most prominent inter-departmental collaborations are between Computer Science and both Biology and Information Sciences. The increase in the assortativity coefficient by department from 2001 to 2003 (shown in Figure 2) can be attributed to a number of factors, including (i) a slight increase in collaborations among Electrical Engineers, (ii) the apperance of novel collaborations among Biologists, and (iii) a substantial drop in collaborations by Computer Scientists with both Biologists and Information Scientists.

In the long run, however, the intra-departmental volume of collaboration among Computer Scientists decreases steadily over time. This decrease, coupled with the growth of a number of inter-departmental collaborations (Computer Science with Electrical and Civil Engineering, as well as Electrical Engineering with Biology and Statistics), is most responsible for the assortativity coefficient trend presented in Figure 2, i.e., the CENS coauthorship network becomes less assortative by department, and thus more inter-disciplinary, over time.

The observed patterns can be interpreted as follows. First the overall presence of intra-departmental collaborations in Computer Science is telling of the nature of research being performed at CENS. The domain of networked sensing emerges historically from computer network research and is thus, normally located as a branch in departments of Computer Science. Sensor network technologies, however, require the design and construction of wireless sensors, and, in turn, interaction between computer sciences and engineering disciplines follows necessarily. This growing incidence of a core set of Electrical Engineering collaboration (both intra- and inter-departmental) is evident in Figure 4. It is interesting to note that inter-departmental collaborations with Electrical Engineers involve not only Computer Science, but also Biology (a major scientific application area for sensor networks) and Statistics (a discipline increasingly required by field scientists to deal with issues related to sensor data cleaning, analysis, and modeling). Finally, it is worth noting that the volume of intra- and inter-departmental collaborations involving the department of Civil Engineering increases over time, possibly reflecting the inception in 2004 of a new application area at CENS involved in the development and application of sensing technologies in urban and social settings.

In sum, CENS, a research center emerged as a sub-domain of Computer Science, has progressively become more inter-disciplinary over time. The increase in inter-disciplinarity can be attributed to CENS' need to develop sensor network technologies (Electrical Engineering), apply and deploy them in field environments (Biologyand Civil Engineering), as well as deal with data analysis issues (Statistics).

3.2.3 Academic position

From Figure 2, it is evident that the discrete assortativity coefficient based on academic position has very little influence on the overall topology of the CENS coauthorship network, compared to the other computed measures. Assortativity by academic position never exceeds a value 0.2 and, very importantly, it remains practically unchanged throughout the period under study. This finding suggests that the network is very weakly assortative with respect to academic position, i.e. coauthorship activities involve the collaboration among scholars of all ranks. It is fair to speculate that this finding is a *de facto* characteristic of scholarly publishing in many academic fields. However, a detailed decomposition of the most prominent academic position pairs can help reveal the specific mixing patterns that influence this assortativity value of the CENS coauthorship network.

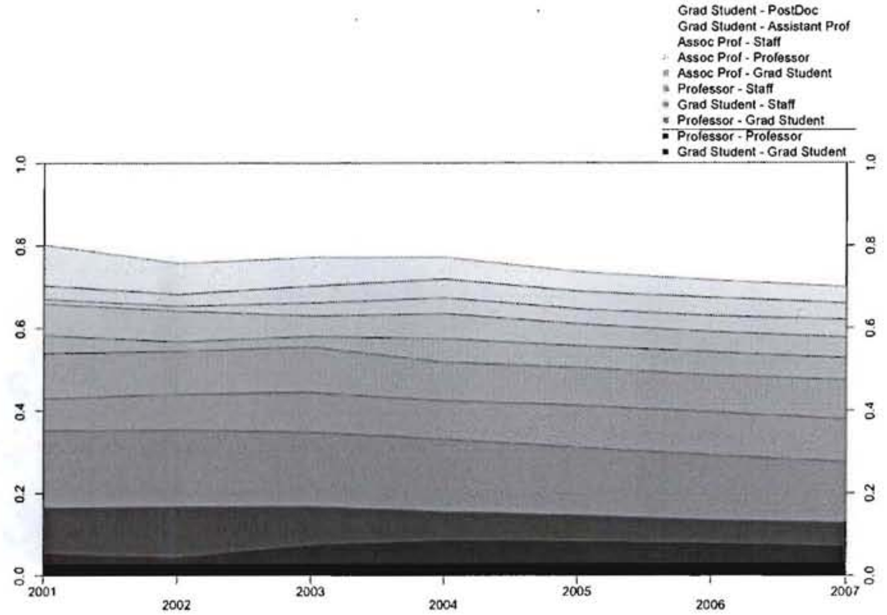

Figure 5: Top ten most recurrent academic position pairs as fraction of total volume of collaboration. Darker polygons at the bottom depict collaborations among pairs of individuals with the same academic position, while lighter polygons depict collaborations between individuals of different academic positions.

Figure 5 presents a stacked plot of the top ten most prominent academic position pairs as fraction of the cumulative, yearly volume of collaboration in the CENS network. The two darker polygons at the bottom depict the position pairs that contribute the most to make the network more assortative — collaborations between both graduate students and full professors. The eight polygons in lighter shades of gray depict the volume of collaborations that make the network less assortative, for they represent coauthorship activity among scholars of different ranks.

The plot of Figure 5 is very variegated and none of the analyzed pairs stands out for volume of collaboration. Moreover, the overall assortativity coefficient by academic position remains constant over time. Yet, certain academic position pairs display some minor fluctuations which we deem interesting. Looking at the darker polygons in Figure 5 we note that collaboration among full professors diminishes slightly from 2001 to 2007, while coauthorship between graduate students increases, in the same period. This result suggests the following research scenario: at its outset, CENS did not include many graduate students; most of the research to "bootstrap" the research center was carried out by faculty members. However, as the center grew in size, more and more graduate students became involved and patterns of collaboration among graduates became more prominent.

This finding is also reflected in the analysis of collaborating pairs of different ranks, i.e., the lighter polygons in the top portion of Figure 5. Again, no major fluctuations emerge from a quick visual analysis of the plot. However, if we look collectively at collaborations of full professors with graduate students, staff members and other professor ranks, we note that they decrease slightly from 2001 to 2007, while collaborations among graduate students, staff, associate and assistant professors increase during the same time period. This finding reinforces the research scenario presented above; it indicates that after a "bootstrap" phase in which scholarly publication was led by high-ranking professors, a younger research network was formed, consisting of graduate students and faculty in the early stage of their careers.

3.2.4 Country of origin

The final discrete assortativity coefficient we analyze in this article is the country of origin of the individuals in the network. Figure 2 shows that this coefficient increases steadily in the first two years and then levels off in later years at a value around $r = 0.35$. What specific intra- and inter-national collaborations may account for such trend? Even if the vast majority of publications analyzed in this study are based on research performed in the United States, it is interesting to explore the tendency of individuals to collaborate with others from their country of origin, even when they are working and living abroad, or in different countries.

We present in Figure 6 a stacked plot of the top ten most prominent country of origin pairs, as fraction of yearly cumulative volume of collaboration. At the network's outset, the vast majority of collaborations is among Americans and between Americans and Indian and Chinese researchers. By year 2003, the picture only changes slightly. More intra-national collaborations appear (China-China and India-India), while inter-national collaborations between USA, India and China drop. These dynamics account for the growth of overall assortativity by country of origin recorded from 2001 to 2003, and visible in Figure 2. By year 2007, the picture becomes more variegated. Intra-national collaborations among American researchers still dominate. However, a number of novel international collaborations emerge, namely between USA and Italy, South Korea, and Iran. Connecting this finding with the observation made in the previous section — that collaboration at CENS was initially "bootstrapped" by faculty and later continued by graduate students and younger faculty — it is possible to infer that as soon as CENS acquired a solid research core of collaboration, by year 2004, it began to attract and involve collaborations by international graduate students, researchers and faculty.

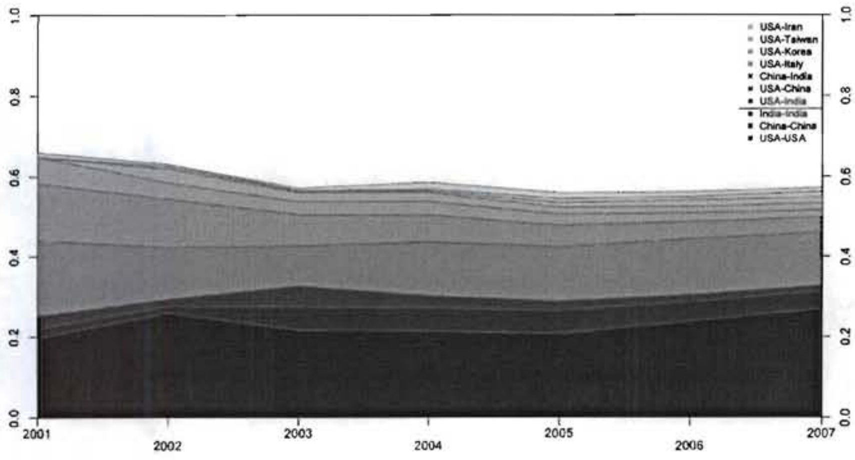

Figure 6: Top ten most recurrent country of origin pairs as fraction of total volume of collaboration. Darker polygons at the bottom depict intra-national collaboration, while lighter polygons depict inter-national collaboration.

4 Conclusion

A great deal of research on scientific collaboration is performed on large-scale networks, constructed from bibliographic data harvested from domain-based and institutional document repositories. While these analyses rely on great quantities of data to study the structure, evolution and similar macroscopic features of scientific collaboration patterns, they often ignore certain contextual and microscopic factors, such as the social and academic arrangements in which collaboration takes place. This is because many available bibliographic datasets contain detailed publication metadata, but very little or no data about the authors writing those publications.

In this article, we perform a longitudinal analysis of the range, configuration and topology of a small network of scientific collaboration over a 7-year period. The network presented here is constructed from the bibliographic record of CENS, a research center involved in the development and application of sensor network technologies. Given the relatively small size of the network ($N = 291$, in its largest year), we were able to manually collect additional metadata for every individual in the network studied. We used these node characteristics to explore the assortative mixing based on academic department, affiliation, position, and country of origin.

Our findings reveal that, in the period under study, the CENS collaboration network: a) becomes more assortative in terms of academic affiliation, i.e. more intra-institutional, b) becomes less assortative in terms of academic department, i.e. more inter-disciplinary, c) is not assortative in terms of academic position, i.e. collaboration patterns are not dependent on researchers' academic positions, and d) is only slightly assortative in terms of country of origin, i.e. the extent of inter-national collaboration decreases slightly over time.

In this article, we interpreted our findings in terms of the specific components that constitute these mixing patterns, finding that a) the increase in intra-institutional collaboration is possibly caused by CENS research consolidating around its headquarter base at UCLA, and finished in 2004; b) the increase in inter-disciplinarity is largely due to the shift a CENS' research agenda, to incorporate new domains, such as civil engineering and urban planning, besides the domains traditionally associated with sensor network research, i.e., computer science and electrical engineering; c) CENS research was initially dominated by collaborations between senior researchers, but as the center matured, it became more diversified and a core research component consisting of collaborating graduate students and young faculty emerged; and d) the volume of international collaboration between USA, India, and China decreased but new smaller international efforts began as the organization matured.

This qualitative explanation of our findings revealed specific small-scale patterns that a quantitative analysis of assortativity alone would have failed to uncover. We speculate that supporting social network analyses with the proposed qualitative investigation of mixing patterns can provide a deeper understanding of the dynamics that shape (and are in turn shaped) by the changing socio-academic landscape in which scientific collaboration takes place.

References

[1] Francisco Acedo, Carmen Barroso, Cristobal Casanueva, and Luis Galan. Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, 43(5):957–983, 2006.

[2] A. L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311 (3-4), 2002.

[3] Jeremy P. Birnholtz. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science & Technology*, 57(13):1758–1770, 2006.

[4] Johan Bollen, Herbert Van de Sompel, Aric Hagberg, Luis Bettencourt, Ryan Chute, Marko A. Rodriguez, and Lyudmila Balakireva. Clickstream data yields high-resolution maps of science. *PLoS ONE*, 4(3):e4803, 03 2009.

[5] Andrea Bonaccorsi and Cinzia Daraio. Age effects in scientific productivity. *Scientometrics*, 58(1):49–90, 2003.

[6] Phillip Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

[7] Christine L. Borgman and Jonathan Furner. Scholarly communication and bibliometrics. *Annual Review of Information Science & Technology*, 36(1):2–72, 2002.

[8] Tibor Braun, Wolfgang Glanzel, and Andras Schubert. Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, 51:499–510(12), July 2001.

[9] Brian V. Carolan. The structure of educational research: The role of multivocality in promoting cohesion in an article interlock network. *Social Networks*, 30(1):69 – 82, 2008.

[10] M. Catanzaro, G. Caldarelli, and L. Pietronero. Social network growth with assortative mixing. *Physica A Statistical Mechanics and its Applications*, 338:119–124, July 2004.

[11] Paul Cilliers. Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management*, 5(2):135–147, 2001.

[12] Blaise Cronin. *The Hand of Science.* Scarecrow Press, 2005.

[13] Blaise Cronin, Debora Shaw, and Kathryn La Barre. A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science & Technology*, 54(9):855–871, 2003.

[14] Trevor Fenner, Mark Levene, and George Loizou. A model for collaboration networks giving rise to a power-law distribution with an exponential cutoff. *Social Networks*, 29(1):70 – 80, 2007.

[15] Jerrold W. Grossman and Patrick D. F. Ion. On a portion of the well-known collaboration graph. *Congressus Numerantium*, 108:129–131, 1995.

[16] Frank Havemann, Michael Heinz, and Hildrun Kretschmer. Collaboration and distances between german immunological institutes - a trend analysis. *Journal of Biomedical Discovery and Collaboration*, 1(1):6, 2006.

[17] A. Hollis. Co-authorship and the output of academic economists. *Labour Economics*, 8:503–530(28), 2001.

[18] Sofía Liberman and Kurt Bernardo Wolf. Bonding number in scientific disciplines. *Social Networks*, 20(3):239 – 246, 1998.

[19] Leah A. Lievrouw, Everett M. Rogers, Charles U. Lowe, and Edward Nadel. Triangulation as a research strategy for identifying invisible colleges among biomedical scientists. *Social Networks*, 9:217–248, 1987.

[20] Xiaoming Liu, Johan Bollen, Michael L. Nelson, and Herbert Van de Sompel. Co-authorship networks in the digital library research community. *Information Processing & Management*, 41(6):1462–1480, 2005.

[21] Lori Lorigo and Fabio Pellacini. Frequency and structure of long distance scholarly collaborations in a physics community. *Journal of the American Society for Information Science & Technology*, 58(10):1497–1502, 2007.

[22] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[23] Salvatore Mele, David Dallman, Jens Vigen, and Joanne Yeomans. Quantitative analysis of the publishing landscape in high-energy physics. *Journal of High Energy Physics*, 12, 2006.

[24] Jason Moody. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69:213–238(26), 2004.

[25] Mark E. J. Newman. Who is the best connected scientist? A study of scientific coauthorship networks. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, editors, *Complex Networks*, pages 337–370. Springer, 2004.

[26] Mark E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132, 2001.

[27] Mark E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20), 2002.

[28] Mark E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.

[29] Mark E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003.

[30] Mark E. J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101 Suppl 1:5200–5205, 2004.

[31] Marko A. Rodriguez and Alberto Pepe. On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics*, 2(3):195–201, 2008.

[32] Joachim Schummer. Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3):425–465, 2004.

[33] Tom A.B. Snijders, Christian E.G. Steglich, and Michael Schweinberger. Modeling the co-evolution of networks and behavior. In *Longitudinal models in the behavioral and related sciences*, Eds. Kees van Montfort, Han Oud and Albert Satorra, Lawrence Erlbaum, 2007.

[34] Eugen Tarnow. Coauthorship in physics. *Science and Engineering Ethics*, 8(2):175–190, 2002.

[35] Marco Tomassini, Leslie Luthi, Mario Giacobini, and William B. Langdon. The structure of the genetic programming collaboration network. *Genetic Programming and Evolvable Machines*, 8(1):97–103, 2007.

[36] Sharon Traweek. *Beamtimes and lifetimes: The world of high energy physicists.* Harvard University Press., Cambridge, MA, 1992.

[37] Caroline S. Wagner and Loet Leydesdorff. Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10):1608 – 1618, 2005.

[38] Stanley Wasserman and Katherine Faust. *Social network analysis.* Cambridge University Press, Cambridge, 1994.

[39] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.