

---

# **AC 2011-2836: LOOSE NETWORKS AND THE COMMUNITY OF ENGINEERING EDUCATION RESEARCH: A DEFINITION BY BIBLIOMETRIC STANDARDS**

## **Johannes Strobel, Purdue University, West Lafayette**

Johannes Strobel is Director of INSPIRE, Institute for P-12 Engineering Research and Learning and Assistant Professor of Engineering Education & Educational Technology at Purdue University. After studying philosophy, religious studies and information science at three universities in Germany, he received his M.Ed. and Ph.D. (2004) in Learning Technologies from the University of Missouri-Columbia, USA. NSF, SSHRC, FQRSC, and several private foundations fund his research. His research and teaching focuses on the intersection between learning, engineering, the social sciences, and technology, particularly sustainability, designing open-ended problem/project-based learning environments, social computing/gaming applications for education, and problem solving in ill-structured/complex domains.

## **Dr David F Radcliffe, Purdue University, West Lafayette**

Dr. David Radcliffe is the Kamyar Haghghi head and Epistemology Professor of Engineering Education in the School of Engineering Education at Purdue. His research focuses on the nature of engineering; engineering habits of mind, how engineering knowledge is created and shared and how it is learned especially outside the classroom. Over the past 20 years he has conducted field research on the practice of engineering design, new product development and innovation in variety of industries, in large and small firms with an emphasis on design thinking, most recently in relation to sustainability. He also studies engineering education as a complex system, and the design and evaluation of next generation learning environments. This research is intrinsically multidisciplinary and draws on methodologies from the humanities, social and behavioral sciences and involves collaboration with anthropologists, learning scientists, librarians, designers and architects.

## **Prashant Rajan, Purdue University, West Lafayette**

Prashant Rajan is a graduate student in the Department of Communication at Purdue University. He is interested in the emergence and sustainability of organizational forms based on the informal networks of collaboration among researchers. Presently, he is working under the guidance of Drs. Johannes Strobel and David Radcliffe.

## **Sadia Nawaz, Purdue University, West Lafayette**

Sadia Nawaz is a graduate student in the School of Electrical and Computer Engineering, Purdue University. She is working with Prof. Strobel and Prof. Radcliffe on data-mining and database management. Her interests also include citation analysis and social networks analysis. Prior to this, she completed her BSc (honors) and MSc in Electrical Engineering from University of Engineering & Technology, Lahore Pakistan.

## **Yi Luo, Purdue University**

Doctoral Student in Learning Design and Technology in Purdue University.

## **Jea H. Choi, Purdue University**

Doctoral student in Learning Design and Technology program in Purdue University.

## **Mapping the domain of engineering education research: Network approaches to bibliometric data collection and analysis**

### **Introduction**

Contemporary scholars have acknowledged the steady progress made toward institutionalization within the domain of engineering education (Johri, 2009; Jisiek, Newswander & Borrego, 2009). Particularly, cyber-based collaborations in the form of virtual organizations (virtual organization) are transforming the engineering education research landscape by equipping researchers in industry, government and university with specialized investigative and educational tools. Virtual organizations serve as networked, informational collaboration resource by hosting document-sharing tools, virtual meeting spaces, online tutorials, and courses frequently enabling access and participation to resources for diverse populations and remotely located communities. Even as virtual organizations emerge as popular delivery mechanisms for implementing and delivering cyber-infrastructure, the factors influencing their success or failure are not very well understood. While many virtual organizations continue to emerge and some virtual organizations are very successful, the factors and the mechanisms contributing to their emergence, success or failure are not well understood. Consequently, the design of these critical pieces of cyber-infrastructure is currently a hit-or-miss process. This paper is part of a larger project addressing this gap by researching the evolution of researcher collaborations in virtual, networked environments.

The overarching goal of the project is to investigate empirically the evolution and emergence of virtual organization to identify the factors which contributed to the growth, evolution and sustainability of collaboration networks and virtual organizations involved in . As a nascent discipline, scholarship within engineering education conveys a sense of ambiguity about its disciplinary identity and affiliations. With regard to the multiple disciplinary influences

on engineering education we identify among scholars an urgent need to define the boundaries of the EER field. Rather than draw rigid boundaries around a yet evolving discipline, we respond to the challenge of measuring the development of engineering education as an interdiscipline by adopting a network perspective to identify and study knowledge communities of relevance to the domain of engineering education research. Following a brief description of what we mean by the network approach, we demonstrate how network approaches can be flexibly adapted to aid the collection and analysis of relatively large data-sets which are characterized by complexity along the semantic, spatial, and temporal dimensions.

In the following sections, we first, outline the process through which co-authorship networks can be generated in order to describe and demarcate, and extend the domain of engineering education in terms of the observed patterns of collaboration among researchers. Specifically, we describe a keyword-based strategy to collect bibliometric data which is relevant to engineering education research (EER). Our aim here is to demonstrate the ability of our data collection effort to locate data which is relevant to, and, representative of the emerging field of research in engineering education. Second, we present a case study based on a data sample collected through our keyword-based search process to explain the dynamics associated with the emergence of research collaboration within the domain of engineering education. The case study comprises a longitudinal (time series) analysis of co-authorship data from the bibliographic records for the Frontiers in Education (FIE) conference. Our analysis explains the FIE in terms of a self-organizing network, which operates in accordance with an internal dynamic of preferential attachment that is reflected in the actions of individual authors.

## **The Network Perspective**

The fundamental characteristic of the network perspective is its focus on developing concepts and seeking information on the *relationships* among entities of interest. These entities and the relations between them can be mapped onto graphs or entered into matrices which depict the structure of relationships (links) between entities (nodes) and, can be analyzed using techniques of matrix algebra and permutation (Wasserman & Faust, 1994). Network analysis refers to the study of the relationships between entities<sup>1</sup> which may exist at multiple levels of analysis including the individual, group, organizational, and interorganizational levels (Wasserman & Faust, 1994). Network analysis provides researchers with a precise vocabulary which allows them to define and measure different characteristics of the physical, political, economic, and social environment in which humans are embedded (Wasserman & Faust, 1994). It expresses these environments and their aspects in terms of the patterns of regularity (or lack thereof) in the relationships between different units which interact with each other. For example, the presence of regular patterns of relationships among social entities (e.g. friendship, marriage) is understood in social network terms as evidence of social structure. Relationships of interest to the network analyst span a gamut of biological, physical, political, social, economic, affective and cognitive ties among many others (Scott, 1991).

Human beings are fundamentally networked organisms. From the networks of interaction between subcellular components and genes which determine susceptibility of an individual (or even a population) to disease to the social networks that influence the spread of diseases such as obesity and influenza through human society, networks determine our health and provide us with a way of understanding human health at multiple levels (Barabási, 2007). The interconnected

---

<sup>1</sup> An entity may refer to social entities such as individuals, groups or, organizations. Alternatively, entities may comprise of human-produced objects that we interact with such as information, technology artifacts and other material and non-material resources. It is important to note that network approaches are focused on the relationships between entities rather than on the entities or their attributes alone.

relationship networks of humans and their effect on the human condition and behavior has received theoretical and empirical attention from researchers in fields as diverse as engineering, biology, sociology, public health, and physics (Albert & Barabási, 2002; Jin, Girvan, & Newman, 2001; Wagner & Fell, 2001).

### **Co-authorship networks**

Advances in information and communication technologies have ensured that our society is more networked today than at any point in history. The “digital footprints” (Pepe & Rodriguez, 2010, p. 84) of collaborations among researchers constitute the products of scholarly exchange in the form of bibliographic records stored on large scientific repositories such as Web of Science<sup>®</sup>. These bibliographic records are a rich source of data on: (i) the productivity of individuals and groups, (ii) the state of knowledge production in disciplines, and, (iii) the collaborative relationships among researchers (Pepe & Rodriguez, 2010; Barabasi et al, 2002). Bibliometrics is an advanced field of research focusing on analyses of bibliographic data in the form of co-authorship, citation, and co-citation patterns (Pepe & Rodriguez, 2010). Co-authorship patterns remain the predominant phenomena of interest for researchers engaged in bibliometric analyses. Research on co-authorship has focused on diverse disciplinary domains including high energy physics (Mele et al, 2006), neuroscience (Braun et al, 2001), nanoscience (Schummer, 2004), economics (Hollis, 2001) and organizational studies (Acedo et al, 2001). In recent years, research on co-authorship patterns has moved in the direction of comparative analyses of co-authorship behavior across disciplines (Newman, 2004), and, investigations on the evolution of scientific collaboration over time and space (Wagner & Leyesdorff, 2005; Barabasi et al, 2002). Network analysis has served a seminal role in providing the theoretical foundations as well as the methodological tools to study co-authorship patterns as well as their implications for scientific

communities (Fenner et al, 2007; Liberman & Wolf 1998; Lievrouw et al. 1987). From a social networks perspective co-authorship patterns arise when two individuals (or, nodes) co-occur in the authorship of a publication (which constitutes the link or, the *edge* connecting the nodes). Thus, a co-authorship network is a set of authors (nodes) who are linked together by the publications (edges) which they co-author. A network of co-authorship can then be expressed in form of a square matrix in which each author occupies a row and a column and the number in each cell corresponds to the presence (or, absence) of a co-authored publication, or, the value (that is, the number of co-authored publications) for a pair of authors in the network.

While academic databases contain large amounts of structured data on co-authorship, extracting and aggregating such data from multiple sources including publications, conference proceedings, books, acknowledgments is a complex task on account of the size of such data (which may run into terabytes) as well as the different structures adopted by different publishing venues and databases. Exercises in collecting bibliometric information are further complicated when the goal is to map an emergent discipline as is our case. Continuing in the vein of the network approach described earlier, we decided to employ techniques of data collection which would not just index a particular field (e.g. known publication names in the engineering and educational disciplines) or a set of fields. Rather we engaged in a recurrent and recursive process of linking with the literature on engineering education research, developing an idea about the locus of the domain formally identified as engineering education research and, reproducing knowledge search practices which extended our notions about the scope and role of engineering education. This process is on-going and therefore recurrent as we continue to engage with the literature within and beyond the formal domain of EER. It is recursive because our experiences in searching, collecting, cleaning, and analyzing data have crystallized certain knowledge

practices. In what follows we describe these knowledge practices which have helped us link with bibliometric data which is highly relevant to the EER domain. We conclude with a case study describing the dynamics of research collaboration within EER.

### **A Keyword-based Scheme to Define EER as a Discipline**

For this paper, the authors have attempted to define engineering education research (EER) as an emerging research area field through a keyword based scheme to define EER - an emerging research area. The methodology of this work is grounded in the EER literature and this article aims to serve as an attempt for a comprehensive and systematic keyword list for EER. The keyword-based scheme refers to a conceptual framework that was developed by the research team we have developed for generating a parsimonious list of keywords which are salient in EER. Keywords can be used to search against either the entire document (i.e. full text search) or the titles and abstract (i.e., title-abstract search) since sometimes the data sources do not provide access to the entire documents. In general, “title-abstract-keyword” search seems to generate more accurate results while “full-text” search offers greater coverage of documents. A detailed description of this approach is provided in the following sections.

### **Data collection**

This project began with the download of engineering education research (EER) articles from ISI WoS over the period 1982 to 2010. Since EER is an emerging interdisciplinary research area it does not have established keywords and their hierarchy and also lacks well established publishing venues for research; thus making it extremely hard to ensure the download of all relevant data for EER. In the first phase of keyword collection, the aim was to be as inclusive in searching for keywords as practical. Thus the research team began with the collection of keywords from a broad range of explicitly engineering education venues and some related

publishing venues; which includes both international and national conferences and journals; were chosen. The complete list of these publishing venues is provided by Nawaz et al (2011). The breadth and extent of the publication venues provides one justification for the validity of the subsequent analysis. An alternate justification is provided in the following text under the heading ‘Source Counts’.

This approach resulted in a collection of 2200 social science keywords from existing literature. Another set of 250 words served as engineering keywords. After going through various refinements and ISI WoS truncation rules the keyword list for social science was shrunk down to 278. Although it is possible to define engineering by using other terms and phrases, but in literature the only keyword used for it is ‘engineering’ itself. So for engineering keyword list it was decided to use only the term engineer\*. Nawaz et al (2011) provide the complete explanation, testing and justification of the rules that shrunk down the keyword lists. Further it was decided to utilize separate databases for these two sets of keywords. The searchable citation databases on WoS comprise two citation indices indexing research published in journals and conference proceedings of the physical sciences and engineering. The remaining three citation indices contain records of research from the social sciences, arts and humanities. By using the proposed methodology 142,981 unique records were downloaded from ISI WoS.

### **Data cleaning**

Using the proposed keyword based methodology detailed in Nawaz et al (2011) 142,981 records were downloaded from WoS. Due to the adopted most inclusive approach some of the downloaded records might not be EER related. To get rid of these extra records data cleaning is needed. As a first step towards data cleaning, the keyword field was extracted from these records. Then hwfc (Hermetic Word Frequency Counter) software was run on these extracted keywords



which converted the keyword phrases into keywords and also provided the rank of these keywords based on their frequency in the complete keyword list. 68,477 unique keywords were obtained as a result. These keywords were taken one at a time and decision was taken whether to mark them as good keywords or bad keywords. Good keywords are the words that (based on their context based usage) may belong to engineering education research articles. Once this manual classification of all the keywords is completed, a code is run on the complete article list (i.e., WoS downloaded records) and hence only those records are parsed out which contain one or more of these keywords in their keyword field. This approach reduced the record count from 142,981 to 50,644 records. Thus the first step towards data cleaning is completed.

### **Validating the keyword search strategy**

It was decided to validate the keyword-based search using a post-hoc analysis in which records of a single source i.e., Frontiers in Education (FIE) were taken for analysis. These FIE records were taken from the complete article list and not just the cleaned version and hence a code was run on the 142,981 records to parse out FIE records. As a result 1869 records were filtered out. Validation analysis is split in two parts as provided here: (a) Source/venue count , (b) Validation.

#### **(a) Source/venue counts**

To begin the source count analysis keyword phrases were extracted from the 1869 FIE records. It was found that only 804 of these records had keywords in them. These keyword phrases were then converted into tokens i.e., one keyword phrase per line. Out of the 3344 tokens, 1244 duplicates were found showing that there is 37.2% keyword re-usage. The remaining 2100 unique keyword phrases were then used towards the later steps. The next step was performed on the cleaned dataset i.e., 50,644 records. A code was run to extract all the records that may

contain one or more of these FIE keyword tokens. Surprisingly the records that (contained one or more of FIE keyword phrases) were parsed out had a count of 48,852. It is one justification of the mentioned keyword based approach and also of our data clean-up methodology in that only the keyword phrases of EER based FIE records were taken. So according to the proposed methodology these keywords are a good representative of the EER community. Since these keywords are providing us 96.46% of the total cleaned up data (i.e., 50,644 records) i.e., the parsed out data using only the FIE keyword phrases is very close to the total cleaned up data, it thus proves the correctness of the data cleaning strategy.

Next, to show the correctness of the proposed keyword based approach, source count analysis is presented. For this analysis, the 'SO' i.e., source field of the 48,852 records was parsed out (where these records are those extracted from the cleaned data version using only the keywords of FIE). Then unique occurrences of the venues were counted. Since there are formatting issues in the 'SO' field of ISI WoS; it was necessary to bring all the sources in an identical format to get rid of the duplicates. Thus some automatization and some manual work was done which reduced the source count from 48,852 to 9042. These venues were then broadly classified into following four categories and a graph was plotted to show their area wise distribution.

- (a) Engineering Education (EE)
- (b) Education (other than EE)
- (c) Engineering (other than EE)
- (d) Others (everything except the above mentioned categories)

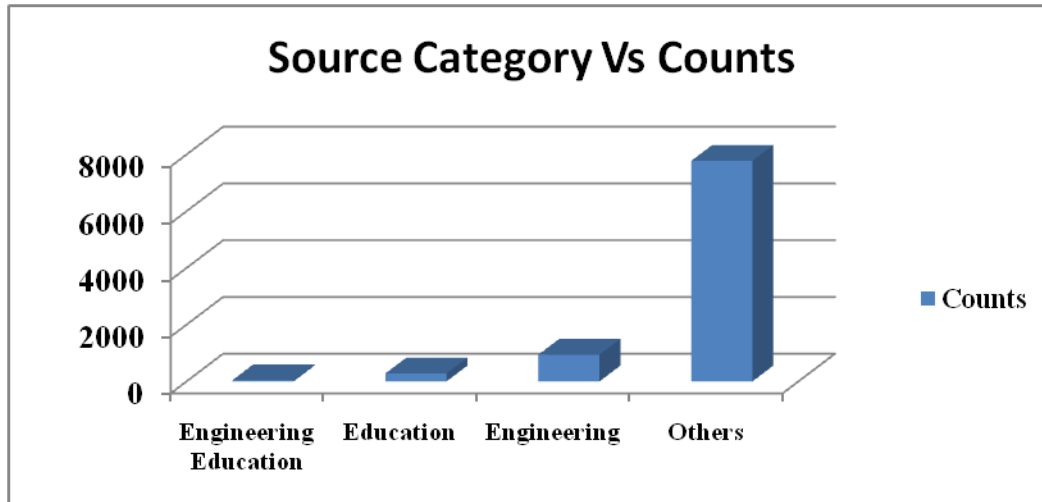


Figure 1. Frequency distribution of sources

This analysis shows that the keywords of a single venue were taken to begin with but as a result we are getting a large number of venues in the outcome which shows the correctness of proposed keyword based search strategy.

It could be argued that the keyword field is empty for some of the articles. As far as data download from ISI WoS is considered, it is not an issue since the keywords are also searched in the title, abstract and subject category and is not restricted to the keyword field itself. However, as far as the proposed data cleaning methodology is considered, the suggested approach is just the first step and further steps are being developed to incorporate other related data (esp. those with empty keyword field) from the downloaded 142,981 records.

**(b) Validation**

We decided to validate the keyword-based search using a post-hoc analysis comparing records downloaded for a single conference – Frontiers in Education with all of the records available for that conference. To establish the validity of our keyword search strategy, we compared the sample of records<sup>2</sup> (FIE\_sample) from the Frontiers in Education (FIE) Conference

<sup>2</sup> A record refers to the bibliographic entry corresponding to a single unique paper in a publication.

for the period 2005 to 2010<sup>3</sup> which were downloaded during the keyword-based search process with all of the FIE records (FIE\_total) available from ISI<sup>®</sup> Web of Science (WoS) over the same duration (2005-2010). The keyword-based search strategy yielded 829 records out of the possible 1265 from FIE for the period 2005 to 2010. In other words the sample generated using the keyword-based strategy covered a little more than 65 % of the total population of records.

<b>Period: 2000 – 2005</b>	<b>Number of records</b>	<b>Percentage</b>
FIE_total (complete record for FIE downloaded from ISI <sup>®</sup> Web of Science)	1265	100 %
FIE_sample (downloaded using keyword-based strategy)	829	65.5 %

Table 1. Share of total records for FIE for the period 2005-2010 which were downloaded using the keyword-based search strategy.

While, the keyword-based strategy yielded nearly two-thirds of the total number of records from FIE, we wanted to establish the relevance of the collected sample by measuring its representativeness vis-à-vis the population of FIE records. To investigate whether we were downloading a sample which was representative of the population, we ranked the top 40 keywords by occurrence for FIE\_sample and FIE\_total (Table 2). If our search strategy was downloading representative records of relevance then we expected to observe the majority of the most frequent keywords co-occurring across FIE\_sample and FIE\_total. Twenty seven out of the top 40 most frequent keywords from FIE\_sample co-occurred in the list of the top 40 keywords from FIE\_total. In other words, the keyword-based strategy yielded records which contained 67.5 % of the top 40 most frequently occurring keywords in FIE. A non-parametric test to

---

<sup>3</sup> While we downloaded records in both our sample and the total population of FIE records for the period 1989 to 2010, our choice of year range was motivated by the fact that keyword fields for FIE records began to be populated in Web of Science after 2005. We report on records for the period 2005-2010 because we were interested in validating our search strategy by comparing keyword occurrences and frequencies.

compute Spearman's rank-correlation coefficient  $\rho$  measured the statistical dependence between the top 40 ranked keywords by occurrence in FIE-sample and FIE\_total. A high, positive value for the rank-correlation coefficient ( $\rho = 0.989$ ) is indicative of the near monotonic trend between the top-ranked keywords in the sample and the total population of FIE records. Finally, to test whether our sample was consistent with population of FIE records we constructed frequency distributions for all those keywords which were located in both FIE\_sample and FIE\_total (Figures 1 and 2). Of the 3127 keywords found in FIE\_total, 690 (~ 22%) were observed to co-occur in FIE\_sample. Pearson's chi-square test was run on the observed (from FIE\_sample) and the expected (from FIE\_total) frequencies of the 690 co-occurring keywords in order to test for the goodness of fit between the distributions of their observed and expected frequencies.<sup>4</sup> A goodness of fit test establishes whether an observed frequency distribution differs from a theoretical (or, in this case expected) distribution. Test statistics for the chi-square test ( $\chi^2 = 200.3$ , with 690 degrees of freedom,  $p = 1$ ) demonstrated that the observed frequencies for the keywords from FIE\_sample were in fact sampled from the expected (FIE\_total) frequencies. In the final section we describe a case study analyzing the dynamics of collaboration among EE researchers who have presented at the FIE conference over the years 1990 to 2010.

---

<sup>4</sup> It is important to note that while records and keywords were observed to co-occur across the sample and the population datasets, the frequency with which a keyword appeared in the sample was different (and typically lower) than the its frequency in the total population of records. The difference in frequencies occurs because the sample does not contain all those records which are indexed by a specific keyword. It is then possible that a particular keyword is ranked lower because it indexes a smaller sample of records. Therefore, to determine whether records downloaded during keyword-based search are a representative sample of the population, we need to ensure that the observed frequencies of keywords (for FIE\_sample records) are sampled from the frequency distribution one would expect to observe in the entire population of records (FIE\_total).

	<b>FIE_sample</b>	<b>FIE_total</b>
1.	engineering education	assessment
2.	assessment	engineering education
3.	education	active learning
4.	retention	education
5.	software engineering	computer science education
6.	active learning	retention
7.	software engineering education	diversity
8.	engineering design	e-learning
9.	engineering	software engineering
10.	diversity	software engineering education
11.	design	engineering design
12.	women in engineering	pedagogy
13.	collaboration	gender
14.	teamwork	computer science
15.	ethics	tablet pc
16.	embedded systems	collaborative learning
17.	tablet pc	distance learning
18.	pedagogy	teamwork
19.	accreditation	ethics
20.	capstone design	globalization
21.	distance learning	collaboration
22.	science	design
23.	abet	embedded systems
24.	service learning	learning
25.	mathematics	service learning
26.	learning styles	accreditation
27.	sustainability	women in engineering
28.	learning	faculty development
29.	experiential learning	distance education
30.	performance	cooperative learning
31.	curriculum	undergraduate research
32.	engineering education research	technology
33.	virtual laboratories	programming
34.	problem based learning	experiential learning
35.	problem solving	curriculum
36.	self-efficacy	project-based learning
37.	professional skills	performance
38.	persistence	science
39.	collaborative learning	mathematics
40.	stem	java

Table 2. Top 40 keywords most frequently occurring keywords in FIE\_sample and FIE\_total

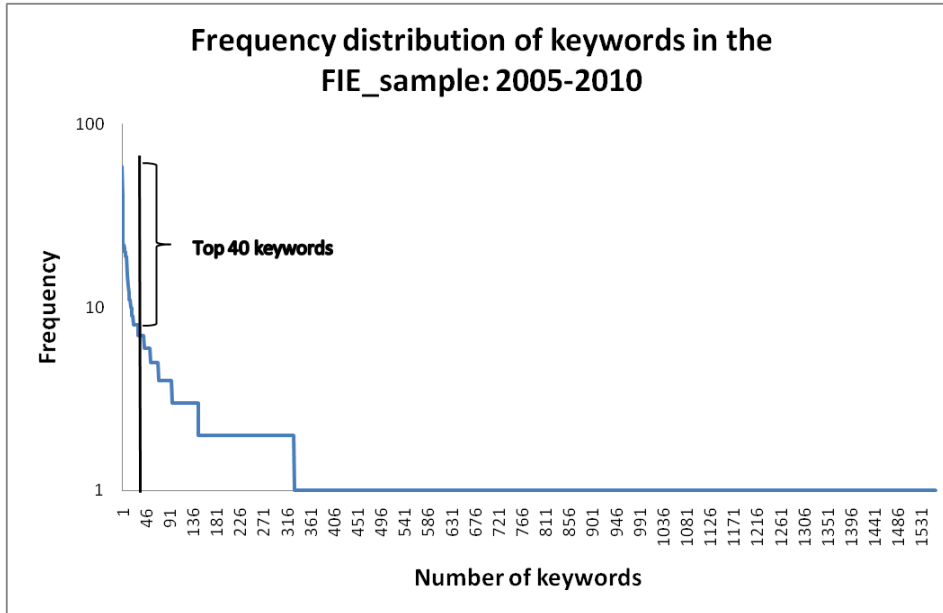


Figure 2. Frequency distribution of keywords from FIE\_sample: 2005 to 2010. Graph plotted on a semi-log scale to clarify location of the top 40 ranking keywords.

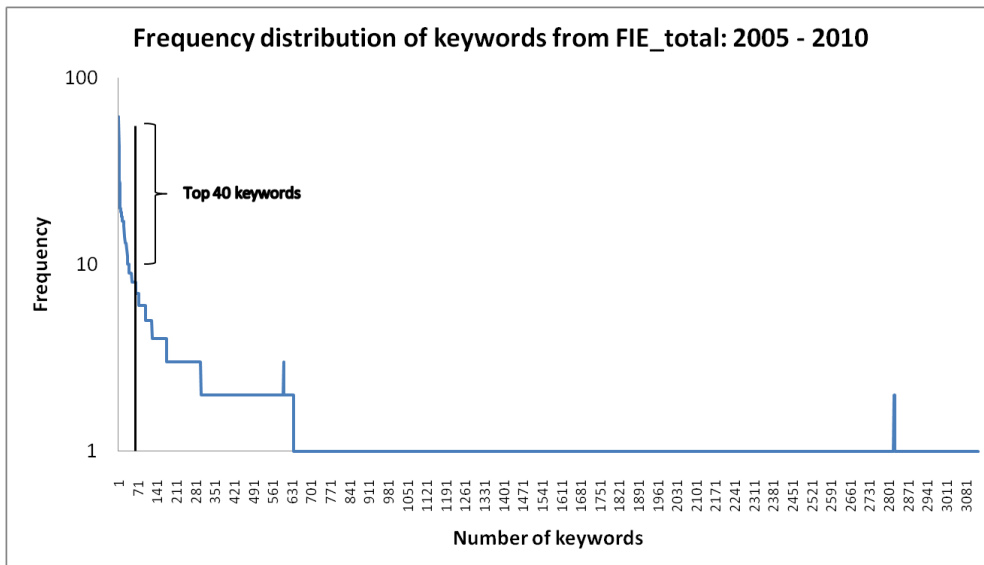


Figure 3. Frequency distribution of keywords from FIE\_total: 2005 to 2010. Graph plotted on a semi-log scale to clarify location of the top 40 ranking keywords.

## **The dynamics of research collaboration in engineering education research: FIE as a case study**

Over the past two decades scholars have developed sophisticated tools to analyze the structure of large networks which exhibit a high degree of interconnectedness among nodes (Jeong et al., 2001; Newman, 2001; Ebel et al., 2002; Dorogovtsev and Mendes 2003; Newman 2004). Such large, interconnected networks possess unique structural signatures such as the presence of clustering among nodes at short length scales and a tendency toward scale-free behavior at the network level (Watts & Strogatz 1998; Barabási & Albert 1999; Jeong et al. 2001). Scientific collaboration networks in particular have been shown to display small average path lengths as well as a high clustering coefficient which vary significantly from their values for random networks of similar sizes (Newman, 2001; 2004). The evolution of such networks occurs through two mechanisms: (a) the addition of new nodes to the network, and, (b) the formation of new links between existing nodes (Jeong et al., 2001). It has also been found that nodes which enjoy a large number of links to others tend to accumulate links at a faster rate than others (Jeong et al., 2001). A colloquial expression for this phenomenon is that the “rich get richer” (Wagner & Leydesdorff, L., 2005). Within a scientific collaboration network, the number of nodes in a network of co-authorship increases as new authors enter the community of co-authors, even as the total number of co-authorship ties increases as existing authors strive to build connections with others (Barabási et al. 2002). Barabási et al (2002) confirmed the tendency among authors in co-authorship networks to display a higher probability of linking with those authors who are in possession of a larger number of co-authorship ties over the network. We explore whether preferential attachment can be observed among research collaborations within the EER domain. The data for this investigation comprises FIE data for the period 1990 to 2010. We further



divided this time period into two subperiods: 1990 – 2000, 2000 – 2010 in order to examine the evolution in the dynamics of co-authorship within FIE over time. We first synthesized three co-authorship matrices – one matrix each for the two sub-periods and a larger matrix containing information on co-authorship over the entire period from 1989 to 2010. Second, we calculated the distribution of co-authorship ties across all of the authors. The number of ties connecting to a node in a network is known as its degree. The distribution of co-authorship ties (or, the degree distribution) is the distribution defining the probability that a node selected at random has  $k$  links. The degree distribution thus, refers to the extent to which an author is well connected within a particular network (Barabási & Albert 1999; Scott 2000; Barabási et al. 2002). A scale-free network is by definition a network for which degree distribution follows a power law for those nodes which exhibit a high degree (have a large number of ties). Thus, the fraction  $P(k)$  of nodes which possess  $k$  links to other nodes is characterized by:

$$P(k) \sim k^{-\gamma}$$

Where  $\gamma$  is a parameter whose values typically range from 2 to 3. Many network in the real world such as networks of actors who have worked with each other, the World Wide Web among other exhibit scale-free behavior (Dorogovtsev & Mendes, 2003). Such networks display long-tailed distributions in addition to containing nodes possessing a degree which is much greater than the average degree for the network. The distribution of degrees influences the topology of the network and creates a hierarchy of nodes in which those with higher degree are linked to others with smaller degree who are in turn linked to others with still smaller degrees (Dorogovtsev & Mendes, 2003). The clustering coefficient for such networks which is simply the ratio of closed (or, clique-like) triangles found in the network to all possible triple tends to decrease as the node degree increases. Thus, nodes with lower degree tend to be embedded in dense local areas of the

network and remain connected to the rest of the network through hubs which possess a much higher.

For the network of co-authorship in FIE, we observe a similar long-tailed degree distribution which may be interpreted as a form of the power-law (Barabási & Albert 1999; Albert & Barabási 2000; Barabási et al 2002). The power-law exponent falls from 3.31 to 2.06 during the periods 1990-2000 and 2000-2010. Value of the exponent for the entire period is 3.46. however, the power-law appears to operate only toward higher degree-end of the distribution. Graphs for all three slices show the presence of a hook toward the lower degrees and a fat-tail toward as the degree rises. Overall, the data suggest that the FIE data follows the scale-free distribution of co-authorship patterns verified earlier for other scientific collaboration networks. Furthermore, the high clustering coefficients are indicative of a high degree of local aggregation which are typical of a network exhibiting “small world” properties (Dorogovtsev & Mendes, 2003). Such networks typically exhibit high clustering coefficients along with short path lengths or distances from one node to another. The column on the extreme right shows the value for the clustering coefficients for randomly generated networks of similar sizes and densities. The three order of magnitude difference between the C-values for their observed and simulated graphs suggests that the properties observed are indeed unique to the network of FIE co-authorship. The slight decrease in C-values over time is indicative of the network becoming relatively less cliquish (or, locally dense) as collaboration patterns begin to even out. Small world networks are highly efficient in transferring information through local exchanges with actors utilizing information and resources that are available to them locally through cliques that they become a part of. The presence of the hook and tail component to the degree distributions is indicative of the environmental constraints on individuals who wish to join such networks. The hook

corresponds to nodes with lower degrees and is associated with the arrival of newcomers into the FIE domain. The slight increase in the value of the exponent in blue regions of the graphs for 1990-2000 and 2001-2010 is indicative of the increase in the numbers of newcomers. The tail-end lies in the high degree section of the graph and represents elite scientists who look to add co-authorship tie rather than look for other authors to link with.

<b>Time period</b>	$\gamma$	$R^2$	<b>Observed clustering coefficient C</b>	<b>Clustering coefficient for randomly generated graphs corresponding to the size and density of the observed network (25 iterations)</b>
1990-2000	3.31	0.79	0.912	0.0015
2001-2010	2.06	0.78	0.805	0.002
1990-2010	3.46	0.83	0.807	0.001

Table 3. Degree distribution and clustering coefficients

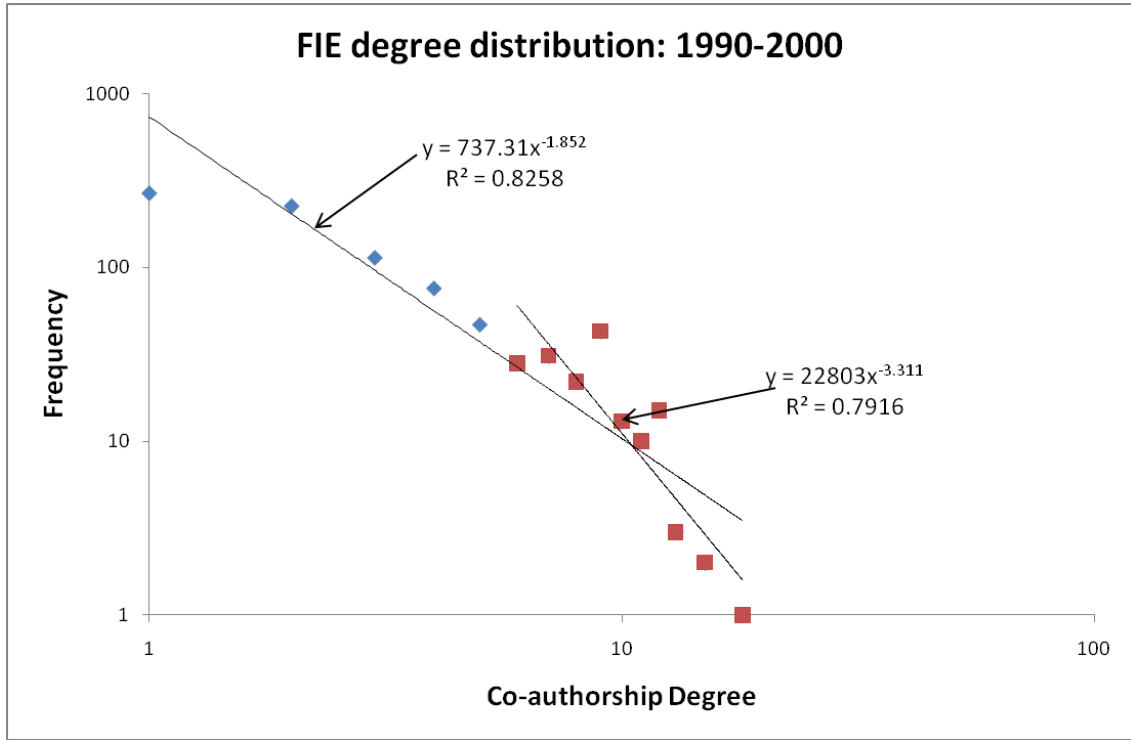


Figure 4. Degree distribution for FIE 1990-2010. Fit to red trace provides the appropriate values for the power-law coefficient  $\gamma$ .

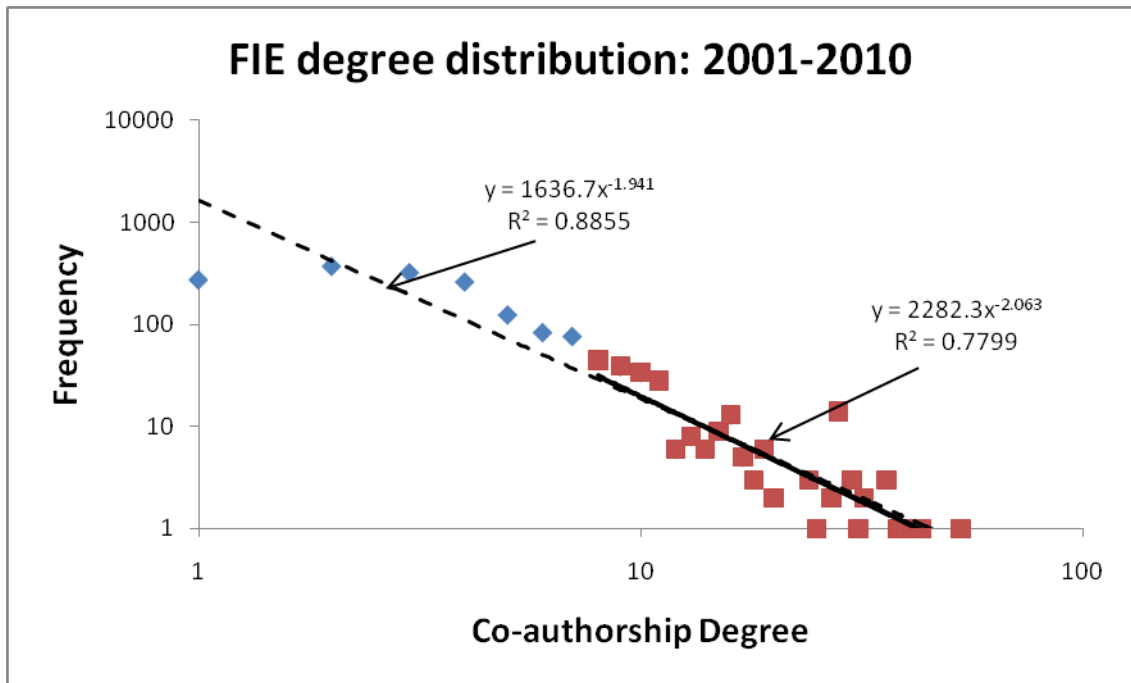


Figure 5. Degree distribution for FIE 2001-2010. Fit to red trace provides the appropriate values for the power-law coefficient  $\gamma$ .

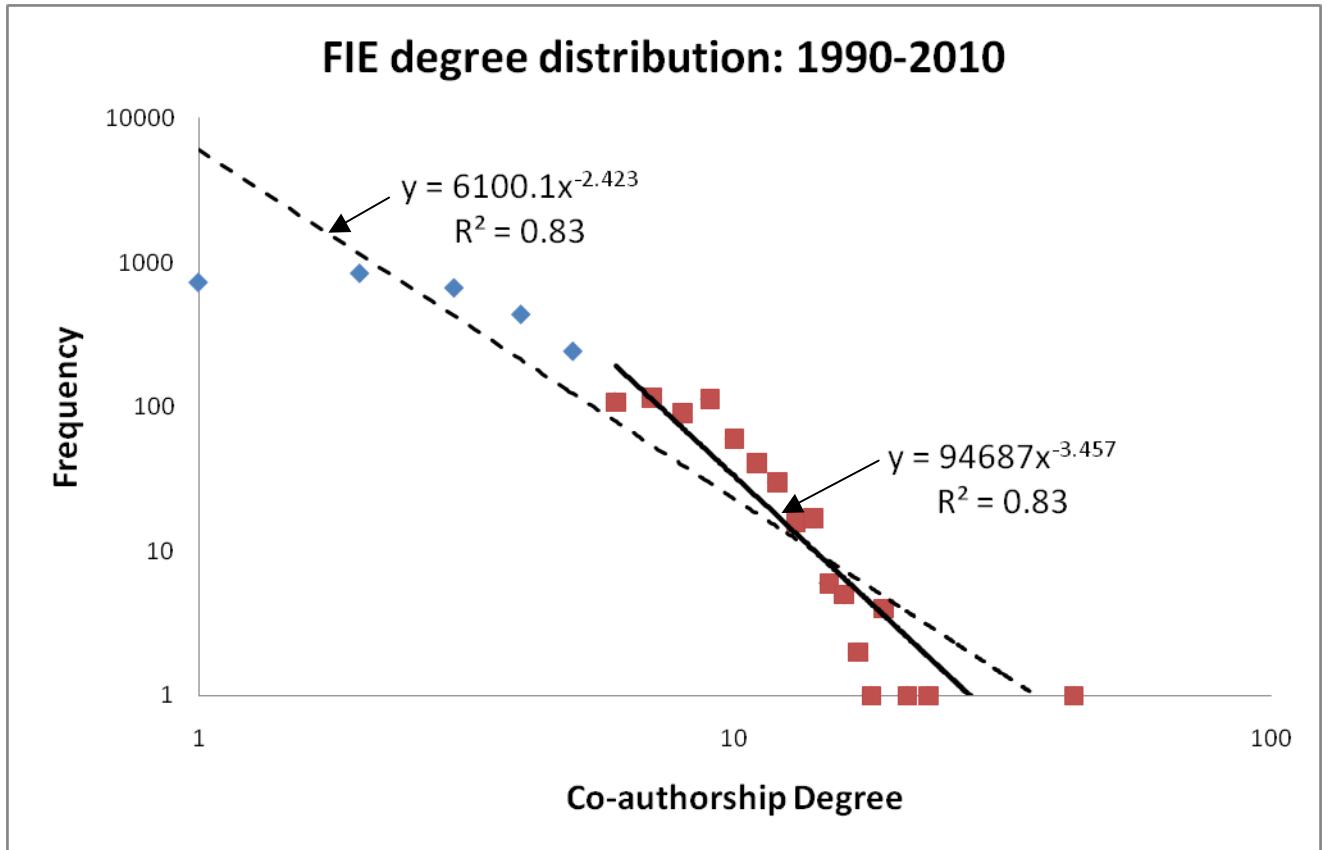


Figure 6. Co-authorship degree distribution in Frontiers in Education (FIE) from 1990 to 2010.

Fit to red trace provides the appropriate values for the power-law coefficient  $\gamma$ .

## References

1. Acedo, F., Barroso, C., Casanueva, C., & Galan, L. (2006). Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, 43(5), 957–983.
2. Jin, E. M., Girvan, M., & Newman, M. E. J. (2001). Structure of growing social networks. *Physical review E*, 64(4), 46132.
3. Barabási, A.L. (2007). Network medicine--from obesity to the "diseasome". *The New England journal of medicine*, 357(4), 404-407.
4. Barabási, A.L., Jeong, H. Neda, Z. Ravasz, E. Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. *Physica A*, 311, 590-614.
5. Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. *Science*, 286, 509-512
6. Braun, T., Glanzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, 51, 499–510
7. Dorogovtsev, S.N., Mendes, J.F.F., 2003, Evolution of Networks From Biological Nets to the Internet and WWW. Oxford University Press, London.
8. Ebel, H., Davidsen, J., Bornholdt, S., 2002. Dynamics of Social Networks. *Complexity*, 8(2), 24-27.
9. Hollis, A. (2001). Co-authorship and the output of academic economists. *Labour Economics*, 8(28), 503–530.
10. Jisiek, B.J., Newswander, L.K., & Borrego, M. (2009). Engineering education research: discipline, community, or field? *Journal of Engineering Education*, 98(1), 32-59.
11. Johri, A. (2010). Creating theoretical insights in Engineering Education. *Journal of Engineering Education*, 99(3), 183-184.
12. Mele, S., Dallman, D., Vigen, J., & Yeomans, J. (2006). Quantitative analysis of the publishing landscape in high-energy physics. *Journal of High Energy Physics*, 12, 1–23.

13. Jeong, H., Néda, A., Barabási, A.L., 2001. Measuring preferential attachment for evolving networks. [http://arxiv.org/PS\\_cache/cond-mat/pdf/0104/0104131v1.pdf](http://arxiv.org/PS_cache/cond-mat/pdf/0104/0104131v1.pdf). Retrieved 22 February 2011.
14. Nawaz, S., Rajan, P., Yu, J.H., Luo, Y., Choi, J.H., Radcliffe, D.F. and Strobel J., "A keyword based scheme to define engineering education research as a field and its members", Accepted for presentation at the IEEE EDUCON 2011, April 4-6, 2011, Amman, Jordan.
15. Newman, M. E. J. 2004. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*, 5200-5205.
16. Newman, M. E. J. 2001. "The structure of scientific collaboration networks". *Proceedings of the National Academy of Sciences* *98*, 404–409.
17. Pepe, A. Rodriguez, M. A. (2010). Collaboration in sensor network research: An in-depth longitudinal analysis of assortative mixing patterns. *Scientometrics*, *84*, 687-701.
18. Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, *59*, 425–465.
19. Scott, J.P., 2000. *Social Network Analysis: A Handbook*. Sage Publications, London.
20. Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings: Biological Sciences*, *268*(1478), 1803-1810.
21. Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, *34*(10), 1608–1618.
22. Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press
23. Albert, R., Barabási, A.L., 2000. Typology of evolving networks: Local events and universality. *Physical Review Letters*, *85*, 5234-5237.
24. Watts, D. J., and Strogatz, S. H., 1998. Collective Dynamics of 'Small-World' Networks. *Nature*, *393*, 440-442.